

UNIVERSITE DE MONTPELLIER 2

Sciences et Techniques

THESE

Présentée pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE DE MONTPELLIER 2

Discipline : Biochimie et biologie moléculaire

Ecole Doctorale SIBAGHE

**(Systèmes Intégrés en Biologie, Agronomie, Géosciences,
Hydrosciences, Environnement)**

Par

Solen ROCHER

**Etudes d'associations pan-génomiques chez la canne à
sucre (*Saccharum* spp.)**

*Analyse de la résistance à l'échaudure des feuilles (*Xanthomonas albilineans*) et
à la maladie de la feuille jaune (Sugarcane Yellow Leaf Virus).*

Présentée le 14 décembre 2011 devant la commission d'examen :

M. QUETIER Francis, Professeur à l'université de Paris XI

Rapporteur

M. GRIVET Laurent, Chercheur (HDR), Syngenta

Rapporteur

Mme DELOURME Régine, Directrice de recherche (HDR), INRA

Examineur

Mme THIS Dominique, Maître de conférences (HDR), SupAgro

Examineur

Mme D'HONT Angélique, Chercheur (HDR), CIRAD

Directeur de thèse

M. HOARAU Jean-Yves, Chercheur, CIRAD

Encadrant de thèse



UNIVERSITE DE MONTPELLIER 2

Sciences et Techniques

THESE

Présentée pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE DE MONTPELLIER 2

Discipline : Biochimie et biologie moléculaire

Ecole Doctorale SIBAGHE

**(Systèmes Intégrés en Biologie, Agronomie, Géosciences,
Hydrosciences, Environnement)**

Par

Solen ROCHER

Etudes d'associations pan-génomiques chez la canne à sucre (*Saccharum* spp.)

*Analyse de la résistance à l'échaudure des feuilles (*Xanthomonas albilineans*) et
à la maladie de la feuille jaune (Sugarcane Yellow Leaf Virus).*

Présentée le 14 décembre 2011 devant la commission d'examen :

M. QUETIER Francis, Professeur à l'université de Paris XI

Rapporteur

M. GRIVET Laurent, Chercheur (HDR), Syngenta

Rapporteur

Mme DELOURME Régine, Directrice de recherche (HDR), INRA

Examineur

Mme THIS Dominique, Maître de conférences (HDR), SupAgro

Examineur

Mme D'HONT Angélique, Chercheur (HDR), CIRAD

Directeur de thèse

M. HOARAU Jean-Yves, Chercheur, CIRAD

Encadrant de thèse

« S'il n'y a pas de solution, c'est qu'il n'y a pas de problème... »

Pr. Shadok

REMERCIEMENTS

Merci tout d'abord à Francis Quetier et Laurent Grivet, pour avoir accepté de juger ce travail en tant que rapporteurs, ainsi que Régine Delourme et Dominique This, pour avoir accepté d'en être examinatrices.

J'adresse également mes remerciements à Angélique D'Hont pour avoir dirigé cette thèse et permis sa mise en place, pour son encadrement pendant mes séjours à Montpellier, et ses conseils toujours judicieux. Je tiens également à remercier Robert Domaingue, chef de l'équipe « Amélioration des plantes à multiplication végétative, qui m'a permis d'effectuer cette thèse dans d'excellentes conditions.

Je tiens tout particulièrement à remercier chaleureusement Jean-Yves Hoarau pour son encadrement, ses conseils, sa patience et sa disponibilité tout au long de ces trois années... avec une pensée toute particulière pour cette dernière période de rédaction. Pour m'avoir permis de découvrir la canne et de saisir sa complexité, pour le temps passé à m'expliquer de mystérieux concepts statistiques, pour m'avoir initiée aux particularités de la recherche aux Antilles, pour m'avoir permis de mener à bien ce travail et bien d'autres choses encore...merci.

Un grand merci à Jean Daugrois, pour avoir partagé avec moi ses connaissances sur la pathologie de la canne, pour son encadrement sur le volet de patho de ma thèse, pour m'avoir initiée à la lecture des symptômes foliaires, et pour sa présence pendant le temps passé en Guadeloupe. J'adresse toute ma gratitude à Samuel Nibouche et Laurent Costet, de l'équipe PVBMT à la réunion, pour les échanges enrichissants que nous avons eus sur les concepts et méthodologie d'association mapping, et pour m'avoir permis de prendre part à cette collaboration fructueuse établie entre nos deux équipes. J'en profite pour remercier Benjamin Fartek et Mathieu Gouy, dont les travaux sur des sujets proches ont été l'occasion de cette collaboration. Je remercie également David Pot, Patrice This, qui ont bien voulu suivre mon travail au cours de ces trois années.

J'adresse mes plus vifs remerciements à tout le personnel de la station de Roujol en Guadeloupe, pour leur accueil chaleureux et leur contribution à ce travail. Je tiens tout particulièrement à remercier les nombreuses personnes qui ont travaillé sur les essais, et notamment Rosiane Boisne-Noc, Jean-Marie Coupan, Sarah Debibakas, Nadia Lubin et Lyonel Toubi, qui ont très largement contribué à la quantité de données de qualité que j'ai utilisées. Merci également à Saturnin Bruyère, Marius Gelabale, Lyonel Toubi, qui se sont abimé les yeux avec moi sur les gels AFLP. Merci à Sylvie, Guy, Nadia, Alice pour les cours de créole, sushis et autres repas.

Merci également à toutes les personnes de l'équipe SEG de Montpellier. En premier lieu, merci à Carine Charron pour ses conseils et ses encouragements, sa présence lors de mes séjours à Montpellier, et pour le temps qu'elle a su trouver pour relire mon manuscrit. Un grand merci également à Catherine Hervouet et Elodie Lacut, pour leur aide précieuse et

joyeuse au labo « P33 », à Elodie, Catherine (encore !) et Cyrille pour avoir successivement ou pas, partagé le bureau, pour l'ambiance le café et les croissants. Merci aussi aux autres membres de l'équipe, ainsi qu'aux membres de l'UMR DAP, pour leur accueil lors de mes nombreux séjours, et parce que grâce à eux ça a toujours été un réel plaisir de me rendre à Montpellier

Ces trois années auront été riches en rencontres, et je tiens à saluer quelques unes des nombreuses personnes qui ont croisé mon chemin sur les hauteurs de Belair, dont la présence a été d'une aide inestimable pour affronter la rudesse de ce climat tropicale et les caps difficiles : Lai Ting (la maison n'a plus jamais eu le même visage après ton départ), Jojo, Elisa, Ben, Fred, Marion Bzz, et les autres, avec un grand merci spécial pour Anne et Marion T. pour les coups de pouce dans les moments difficiles, et parce que finalement, ma thèse, ça revient à chercher du tout petit dans du trop grand. Merci Mylène pour Sun...j'espère bien avoir le temps maintenant. Merci aussi à Franz, JP, Francianne, José,...zot ka fè enmé mwen péyi la on pil !!! A beaucoup d'autres dont j'ai croisé la route pendant ces trois années mémorables. Sans oublier ceux qui même de loin sont toujours là : So, Eole, Marionette, Marion, Gaël...heureusement qu'il y a le tunnel !

Enfin, je tiens tout particulièrement à remercier ma famille, mes parents, pour avoir toujours été là et m'avoir toujours soutenue, mes grands-parents pour être un exemple en bien des choses. Merci à Nana pour le modèle, l'inspiration et le professeur Shadoko, à Bren pour ton grand cœur et parce que je crois en toi. Merci à Pauline et Kevin parce que quand même, à trois c'est plus sympa. Merci à Lyliane pour la cornaline et les baguettes magiques... Merci à ma famille, parce qu'elle est tout simplement.

SOMMAIRE

PREMIERE PARTIE : INTRODUCTION.....	1
I.1 LA CANNE A SUCRE, ELEMENTS DE BIBLIOGRAPHIE GENERAUX.....	2
<i>I.1.1 Histoire et diffusion de la culture de la canne à sucre.....</i>	<i>2</i>
<i>I.1.2 Description morphologique de la plante et cycles de culture.....</i>	<i>3</i>
<i>I.1.3 Valorisation économique.....</i>	<i>5</i>
I.2 TAXONOMIE ET EVOLUTION DE LA CANNE A SUCRE.....	6
<i>I.2.1 Les espèces du genre Saccharum.....</i>	<i>7</i>
a. Les espèces sauvages.....	7
b. Les espèces domestiquées.....	8
<i>I.2.2 Evolution dans le « complexe Saccharum ».....</i>	<i>9</i>
I.3 AMELIORATION GENETIQUE DE LA CANNE A SUCRE	9
<i>I.3.1 Floraison et hybridation.....</i>	<i>11</i>
<i>I.3.2 Critères de sélection.....</i>	<i>12</i>
<i>I.3.3 Principe des schémas de sélection</i>	<i>13</i>
I.4 LES MALADIES DE LA CANNE A SUCRE.....	14
<i>I.4.1 Panorama général.....</i>	<i>14</i>
a. Importance de la sélection contre les maladies	14
b. Abondance de maladies systémiques potentiellement dommageables.....	14
<i>I.4.2 L'échaudure des feuilles.....</i>	<i>15</i>
a. Impact économique, agent causal et gamme d'hôtes	15
b. Expression des symptômes	16
c. Diversité génétique du pathogène.....	18
d. Sources de résistance de l'hôte et nature du déterminisme	19
<i>I.4.3 La feuille jaune.....</i>	<i>20</i>
a. Impact économique, agent causal	20
b. Expression des symptômes	21
c. Transmission de la maladie.....	22
d. Diversité génétique du pathogène.....	23
e. Sources de résistance chez l'hôte et nature du déterminisme.....	24
I.5 GENOME DE LA CANNE	25
<i>I.5.1 Génome des variétés modernes.....</i>	<i>25</i>
<i>I.5.2 Cartographie génétique.....</i>	<i>27</i>
a. Cartes génétiques.....	27
b. Etudes QTL	28
I.6 GENETIQUE D'ASSOCIATION.....	30
<i>I.6.1 Avantages de la génétique d'association</i>	<i>31</i>
<i>I.6.2 Paramètres pris en compte dans les études d'associations.....</i>	<i>31</i>
a. Le déséquilibre de liaison	31
b. Structure et Kinship	32
<i>I.6.3 Etudes d'associations chez la canne à sucre.....</i>	<i>34</i>
a. Structure du DL	34
b. Associations marqueurs-caractères.....	35
I.7 PRESENTATION DES OBJECTIFS DE LA THESE	38
DEUXIEME PARTIE : ETUDE D'ASSOCIATIONS RELATIVE A LA RESISTANCE DE LA CANNE A SUCRE A LA CONTAMINATION AERIENNE PAR XANTHOMONAS ALBILINEANS.....	40
I.1 INTRODUCTION.....	41

I.2 PUBLICATION ASSOCIÉE	41
<i>Abstract</i>	42
<i>Introduction</i>	43
<i>Materials and methods</i>	45
Cultivar panel and field experimentations	45
Genotyping	46
Statistic analysis	46
Assessment of the genetic value of the cultivar panel for leaf-scald resistance.....	46
Panel genetic structure	47
Marker-trait associations	48
Proportion of disease resistance variance explained by markers	49
<i>Results</i>	49
Leaf-scald (LS) assessment of the cultivar panel.....	49
Marker-trait associations within trials	50
Consistency of genetic effect across trials.....	50
Leaf-scald resistance variation explained by markers	51
<i>Discussion</i>	51
<i>Figure and tables</i>	55
<i>References</i>	58
I.3 DONNEES COMPLEMENTAIRES	60
I.3.1 Comparaison des différents modèles d'associations.....	60
I.3.2 Détermination de l'origine ancestrale spécifique des marqueurs.....	61
a. Etude de diversité pour l'identification sécurisée de deux panels de clones ancestraux	62
b. Origine spécifique des marqueurs DArT	63
c. Origine spécifique des marqueurs AFLP	64
d. Origine des marqueurs associés au phénotype agronomique	66

TROISIEME PARTIE : ETUDE D'ASSOCIATION RELATIVE A LA RESISTANCE DE LA CANNE A SUCRE A LA FEUILLE JAUNE, MALADIE CAUSEE PAR LE SUGARCANE YELLOW LEAF VIRUS..... 68

I.1 INTRODUCTION	69
I.2 PUBLICATION ASSOCIEE	69
<i>Abstract</i>	70
<i>Introduction</i>	71
<i>Material and methods</i>	72
Plant material.....	72
Field trials.....	73
SCYLV incidence assessment	73
Aphid incidence recording.....	74
Genotyping	74
<i>Statistical analysis</i>	75
Prediction of genotypic value of cultivars	75
Panel genetic structure.....	76
Marker-trait associations	76
<i>Results</i>	77
SCYLV and aphid incidence	77
Panel structure	78
Marker trait-associations	78
<i>Discussion</i>	80
<i>References</i>	82
<i>Tables</i>	85
<i>Figure captions</i>	87

<i>Supplementary tables</i>	89
I.3 DONNEES COMPLEMENTAIRES	92
<i>I.3.1 Comparaison des modèles d'associations utilisés dans les deux populations étudiées</i>	92
<i>I.3.2 Variation phénotypique expliquée par les marqueurs</i>	93
<i>I.3.3 Détermination de l'origine ancestrale spécifique des marqueurs</i>	96
QUATRIEME PARTIE : CONCLUSION, PERSPECTIVES	97
I.1 RAPPEL DU CADRE DE L'ETUDE	98
I.2 RESULTATS	99
<i>I.2.1 Résultats généraux des modèles d'associations</i>	99
<i>I.2.2 Variations agronomiques expliquées par le marquage et identification d'allèles d'intérêt prioritaires</i>	100
<i>I.2.3 Comparaison des associations détectées dans les deux panels</i>	101
I.3 PERSPECTIVES	102
<i>I.3.1 Un retour à l'étude de ressources génétiques élit</i>	102
<i>I.3.2 A la recherche de nouvelles techniques de marquage</i>	103
REFERENCES BIBLIOGRAPHIQUES	105
ANNEXE I	121

LISTE DES FIGURES

Figure 1 : Plant de canne à sucre issu de la germination d'un bourgeon axillaire d'une bouture.....	3
Figure 2 : Morphologie de la canne.	4
Figure 3 : Cycles pluriannuels de culture de la canne à sucre	4
Figure 4 : Position taxonomique du genre <i>Saccharum</i>	6
Figure 5 : Evolution et domestication de la canne à sucre	9
Figure 6 : Floraison et hybridation.....	11
Figure 7 : Représentation schématique du processus de sélection clonale de la canne à sucre pratiqué en Guadeloupe.....	13
Figure 8 : Echaudure des feuilles de canne à sucre.....	17
Figure 9 : Cycle infectieux de <i>X. albilineans</i>	18
Figure 10 : Organisation du génome du SCYLV et localisation des fragments RT-PCR permettant de différencier les génotypes CUB, BRA-PER et REU.....	21
Figure 11 : Symptômes de la maladie de la feuille jaune causée par le SCYLV.....	21
Figure 12 : Empreintes sérologiques (TBIA – Tissue Blot Immuno Assay) de nervure centrale de feuilles de canne révélant la présence du SCYLV dans les cellules compagnes du phloème.	22
Figure 13 : Structure du génome du cultivar R570 (2n = ca 115).....	26
Figure 14 : Représentation schématique d'une classe d'homéologie caractéristique d'un cultivar de canne à sucre..	27
Figure 15 : Comparaison entre (a) un exercice de détection de QTL dans une population en ségrégation et (b) un exercice de détection d'associations au niveau pan-génomique..	31
Figure 16 : Représentation schématique des différents types de populations rencontrées en génétique d'association.....	33
Figure 17 : Décroissance du déséquilibre de liaison (DL) le long du génome dans une population de cultivars de canne à sucre (<i>Saccharum</i> spp.).....	35
Figure 18 : Distribution cumulée des p-value observées (P_{obs}) par rapport à une loi de probabilité uniforme, des différents modèles d'associations utilisés (Naïf, Q, K, Q+K), concernant la résistance à la propogation aérienne de l'échaudure des feuilles.....	60
Figure 19 : Nombre d'associations détectées par chacun des modèles (Q, K et Q+K) pour la résistance à la contamination aérienne à l'échaudure des feuilles, dans les deux essais, à différentes P-values nominales.....	60
Figure 20 : Arbres de diversité représentant la distance euclidienne moyenne entre clones ancestraux, calculée à l'aide de 625 marqueurs DArT codés dans la population d'association de Guadeloupe.....	63
Figure 21 : Distribution cumulée des p-value observées (P_{obs}) par rapport à une loi de probabilité uniforme, des différents modèles d'associations utilisés (Naïf, Q, K, Q+K), concernant la résistance à la feuille jaune.	92
Figure 22: Nombre d'associations détectées par chacun des modèles (Q, K et Q+K) pour la résistance au SCYLV, dans les deux essais, à différentes P-values nominales.	92
Figure 23: Comparaison de la structuration des deux populations (Guadeloupe et Réunion).93	

LISTE DES TABLEAUX

Tableau 1 : Production annuelle moyenne de sucre de canne dans le monde entre 1999 et 2009 et principaux pays producteurs.....	3
Tableau 2 : Maladies de la canne recensées en Guadeloupe	15
Tableau 3 : Cartes génétiques de la canne à sucre.	27
Tableau 4 : Principales études QTL réalisées chez la canne à sucre.	28
Tableau 5 : Quelques études d'associations réalisées chez plusieurs espèces végétales.....	29
Tableau 5 (suite).	30
Tableau 6 : Etudes d'association réalisées chez la canne à sucre.....	36
Tableau 7 : Liste et provenance des clones ancestraux <i>S. officinarum</i> et <i>S. spontaneum</i> utilisés pour la détermination de l'origine ancestrale spécifique des marqueurs DArT.	62
Tableau 8 : Origine ancestrale des marqueurs associés à la propagation aérienne de l'échaudure des feuilles ($P_{\text{GWER}} < 0,05$).....	64
Tableau 9 : Résultats de régressions linéaires multiples réalisées sur les marqueurs significatifs ($P_{\text{GWER}} < 0.05$) indépendants, pour l'incidence du SCYLV à la Réunion et en Guadeloupe.....	95
Tableau 10 : Origine ancestrale des marqueurs associés à l'incidence du SCYLV à la Réunion et en Guadeloupe ($P_{\text{GWER}} < 0,05$)	96

Première partie :
Introduction

I.1 La canne à sucre, éléments de bibliographie généraux

I.1.1 Histoire et diffusion de la culture de la canne à sucre

Base de la production mondiale de sucre, la canne à sucre est cultivée depuis la préhistoire pour le saccharose qu'elle accumule dans les tiges. La canne dite « noble », *Saccharum officinarum*, est originaire de l'actuelle région de Papouasie Nouvelle-Guinée. Cette espèce riche en sucre aurait été domestiquée vers 8000 avant JC à partir de l'espèce sauvage *S. robustum*, et progressivement diffusée vers les îles du Pacifique, puis la Chine et l'Inde à partir du XIIe siècle avant notre ère. Des croisements interspécifiques naturels entre *S. officinarum* et *S. spontaneum*, une espèce sauvage originaire d'Asie, sont à l'origine de l'émergence de deux espèces : *S. sinense* en Chine et *S. barberi* en Inde, et ont permis la diffusion de la canne dans les régions subtropicales (Daniels and Roach, 1987). L'extraction et l'exploitation industrielle du sucre semblent avoir débuté en Asie. Bien que cette origine reste incertaine, le mot sucre, issu du sanskrit « sarkara » qui signifie grain de sable, témoigne de l'existence de ces techniques en Inde vers 3000 ans avant JC (Meyer, 1989). La canne a atteint l'est de la Méditerranée dès 500 ans avant J.C. grâce au commerce avec la Perse. Alexandre le Grand mentionne ce « roseau qui donne du miel sans le secours des abeilles », qu'il ramène de sa campagne en Inde vers 326 avant notre ère. Les Arabes la diffusèrent jusqu'à l'extrême ouest du bassin méditerranéen ainsi que dans les îles de la Méditerranée à partir du VIIIe siècle, et améliorèrent les techniques de culture de la canne et d'extraction du sucre.

Au XVème siècle, la canne à sucre fut introduite par les espagnols et les portugais dans leurs possessions africaines. C'est à la même période qu'elle a été introduite dans le Nouveau Monde, aux Antilles où elle fut amenée par Christophe Colomb lors de son second voyage en 1493, ainsi qu'en Amérique latine. Les premières plantations dans les Antilles n'apparaîtront cependant que tardivement, vers 1643, suite à l'échec de la culture de tabac. Favorisée par un climat propice, la canne à sucre s'est révélée très rentable, mais également très exigeante en main d'œuvre. Les conditions de travail exécrables et les maladies amenées par les colons ayant rapidement eu raison des populations indigènes recrutées par la force et des flux de « volontaires » venus d'Europe, la main d'œuvre a progressivement été remplacée par des esclaves venus d'Afrique. Ce fut l'origine du commerce triangulaire basé sur un

Tableau 1: Production annuelle moyenne de sucre de canne dans le monde entre 1999 et 2009 et principaux pays producteurs (source FAO, <http://faostat.fao.org>).

Pays	Production annuelle moyenne entre 1999 et 2009 (tonnes)	Rang
Brésil	449 988 545	1
Inde	293 733 000	2
Chine	94 222 195	3
Thaïlande	59 558 427	4
Pakistan	50 878 545	5
Mexique	48 638 345	6
Colombie	37 659 018	7
Australie	35 165 218	8
États-Unis	30 687 755	9
Indonésie	25 969 545	10
Philippines	23 635 155	11
Argentine	23 342 727	12
Cuba	22 569 091	13
Afrique du Sud	21 004 227	14
Guatemala	18 066 518	15
Égypte	16 225 582	16
Viet Nam	16 138 618	17
-----	-----	-----
Réunion	1 846 971	46
Guadeloupe	776 196	61
Martinique	206 737	76
Monde	1 414 880 282	

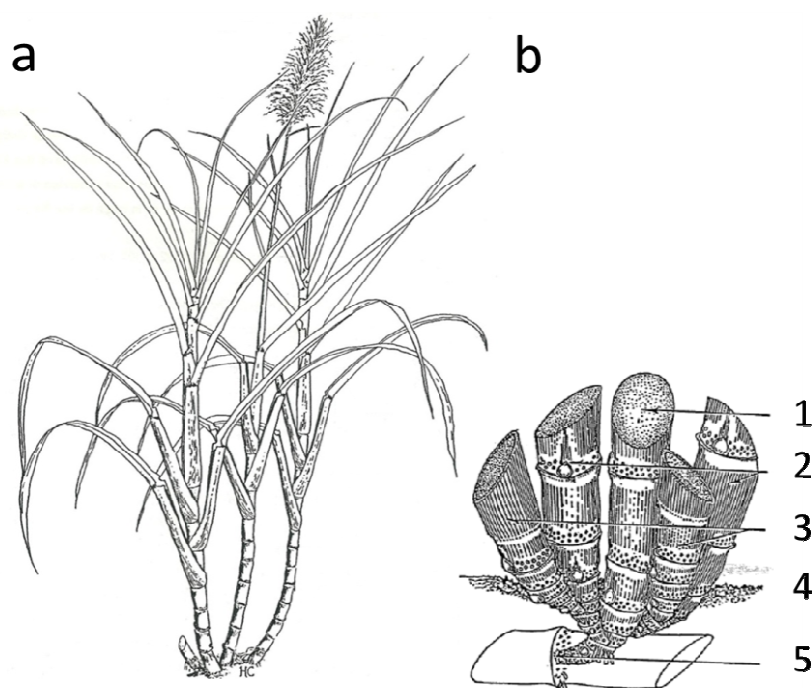


Figure 1 : Plant de canne à sucre issu de la germination d'un bourgeon axillaire d'une bouture.

a) Anatomie générale d'une touffe de canne à sucre (d'après Chaillet, 2000).

b) Ramifications souterraines du tallage : (1) tige primaire, (2) tige secondaire, (3) tige tertiaire, (4) niveau du sol, (5) point d'attache de la bouture originelle (d'après Artschwager and Brandes, 1958).

système esclavagiste, qui a façonné l'histoire des Antilles pendant près de 4 siècles et a profondément marqué la société et la culture antillaise. On estime que 11 millions de personnes ont été déportées vers le Nouveau Monde pendant cette période (Habib, 2008).

Le sucre devient un produit de consommation courante dès les années 1780. La France est à cette époque le principal consommateur et exportateur de sucre, grâce à cet « or blanc » qu'elle importe de ses colonies. A partir de la fin du XVIII^{ème} siècle, le blocus consécutif à la guerre avec l'Angleterre limite les communications avec les colonies et paralyse le transport des produits en provenance d'outre-Atlantique, occasionnant une flambée des prix du sucre qui lève les réticences sur la production de sucre de betterave. Celle-ci prendra progressivement une part de plus en plus importante dans l'approvisionnement en sucre de l'Europe, affaiblissant le lobby sucrier colonial. La première guerre mondiale porta cependant un coup d'arrêt à la production de sucre de betterave, et la canne à sucre reprit une place de choix dans la production mondiale. Elle représente aujourd'hui encore près de 70% de la production mondiale de sucre, les 30 % restants étant principalement extraits de la betterave. La consommation mondiale de sucre est en augmentation constante depuis le début du XX^{ème} siècle, notamment du fait de la demande croissante des pays émergents. La zone de prédilection pour la culture de la canne à sucre se situe en zone tropicale et subtropicale entre 35° de latitude nord et 30° de latitude sud, à des altitudes très variables (0 à 1600 m), sous des climats présentant des pluviométries diverses (500 à 2500 mm) (Blume, 1985). L'air de distribution de la canne, qui est cultivée dans plus de 100 pays à travers le monde et dans des environnements variés, témoigne de sa plasticité phénotypique. Les principaux pays producteurs sont le Brésil, l'Inde et la Chine (**Tableau 1**, source FAO 2009, <http://faostat.fao.org>).

I.1.2 Description morphologique de la plante et cycles de culture

La canne à sucre est une plante semi-pérenne cultivée pendant plusieurs cycles annuels successifs. La plante est un clone, composé d'un nombre variable de tiges (ou talles) rassemblées en touffes (**Figure 1a**). Lors du premier cycle de culture succédant à la plantation, les tiges sont issues de ramifications souterraines de la tige principale qui provient de la germination de l'œilleton de la bouture (**Figure 1b**). Après récolte, les tiges qui

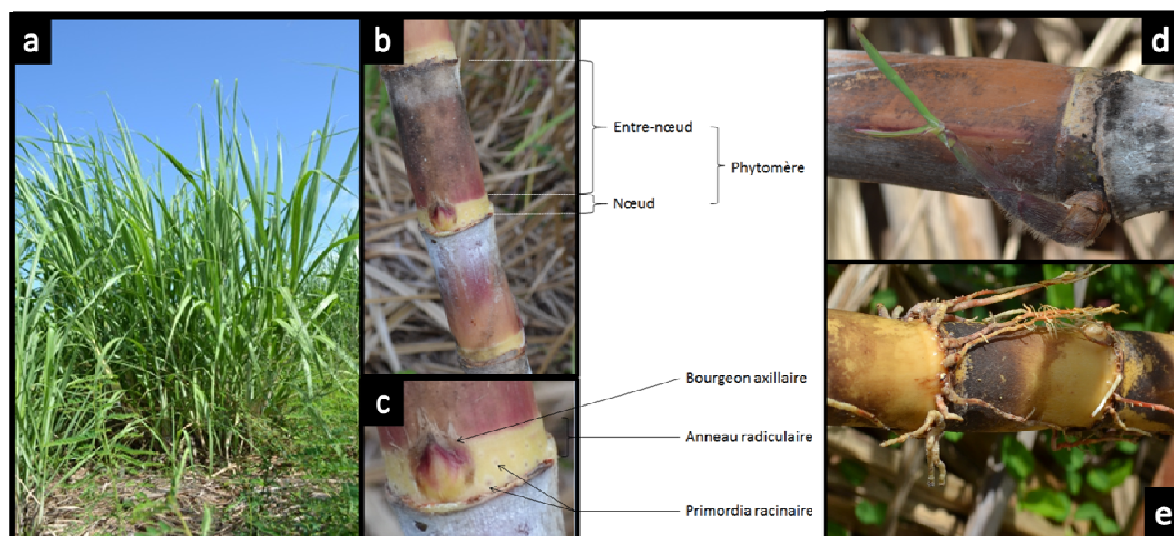


Figure 2 : Morphologie de la canne. a) Touffe de canne à sucre en phase d'élongation b) Détail d'une tige c) Détail d'un nœud d) Germination d'un bourgeon axillaire, évoluant vers la formation d'une tige e) Différenciation de racines à partir des anneaux radiculaires (Photos S. Rocher)

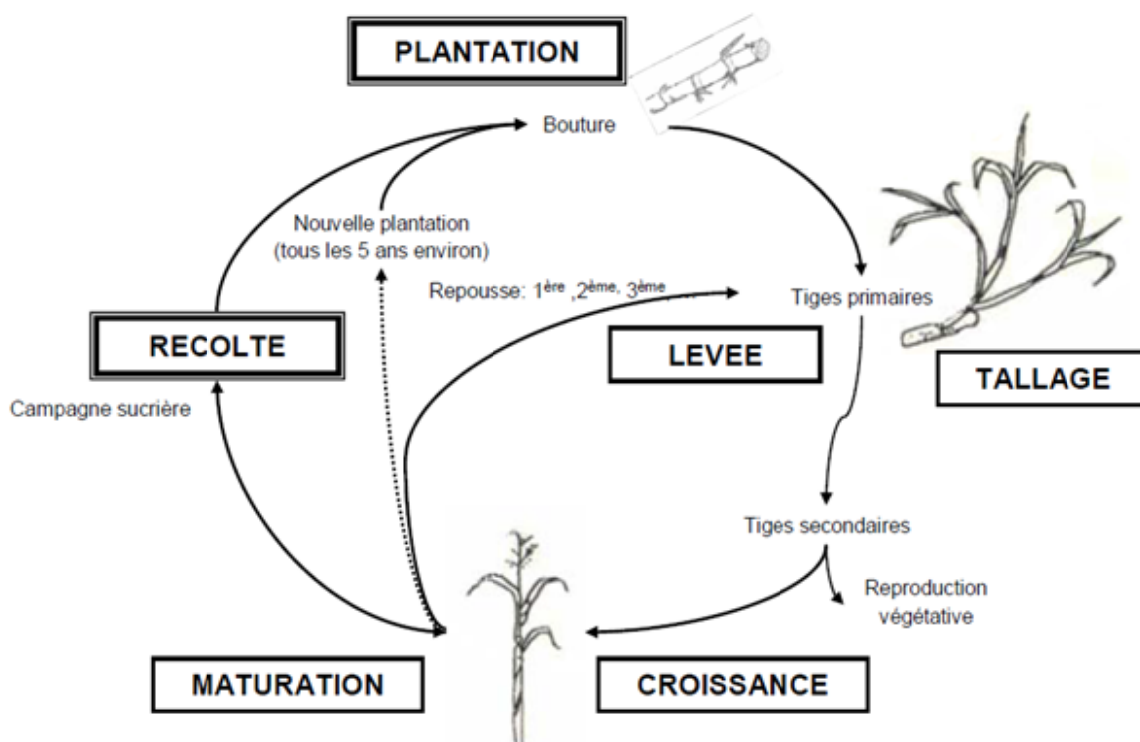


Figure 3 : Cycles pluriannuels de culture de la canne à sucre (Repris de Edon, 2008).

apparaissent en repousse proviennent de la germination du plateau de tallage souterrain formé lors du premier cycle de culture. Chaque tige correspond à une succession d'unités végétatives, les phytomères. Chaque phytomère est constitué d'un nœud et d'un entre-nœud auquel est accrochée une feuille constituée d'une gaine enveloppante et d'un limbe (**Figures 1 et 2**). Les nœuds sont constitués d'un bourgeon axillaire, également appelé œilleton, et de primordia racinaires organisés en un anneau radiculaire. Mis en terre, au contact de l'eau, chaque œilleton est susceptible de germer et de produire une tige. Les boutures, mode de propagation privilégié des cannes cultivées, sont des fragments de tiges constituées de quelques phytomères. Le système racinaire se développe dans un premier temps autour de l'anneau radiculaire, formant des racines primaires superficielles et ramifiées, puis des racines de soutien plus profondes et des racines « cordons » qui servent à l'ancrage de la plante dans le sol (Van Dillewijn, 1960).

Une grande partie des caractères morphologiques, comme la densité du tallage, le port de la tige et des feuilles, la forme de la gaine, la taille et la couleur des entre-nœuds, la présence de cires, *etc.*, peuvent varier en fonction de l'environnement, du génotype et du cycle de culture considéré. De nombreux caractères (morphologie végétative, performances agronomiques, inflorescences, épiderme) sont utilisés de manière simultanée par les sélectionneurs comme clés de détermination et de différenciation des variétés. Cependant, les caractères sont plus ou moins stables d'un environnement à l'autre ou entre cycles différents. Le diamètre et la couleur des tiges, ainsi que certains caractères de l'inflorescence, sont les principaux caractères héréditaires. L'utilisation de multiples critères discriminants ne suffit pas toujours à différencier entre eux de nombreux cultivars. L'utilisation de marqueurs moléculaires est souvent nécessaire pour obtenir des identifications sécurisées ainsi que pour éclaircir des informations d'apparentement ou de phylogénie (Piperidis *et al.*, 2010).

Le cycle annuel de culture de la canne à sucre peut être divisé en plusieurs phases : la plantation, la levée, le tallage et la croissance des tiges, la floraison, la maturation et la récolte (**Figure 3**). La plantation qui peut avoir lieu tous les 4 à 10 ans selon les régions, se fait à l'aide de boutures de tiges. Le tallage, correspondant à l'émission de nouvelles tiges, et la croissance des tiges se font de manière simultanée. La croissance est généralement active pendant les 6 premiers mois du cycle de culture, quand les conditions d'humidité et de température sont favorables. La maturation de la canne, c'est-à-dire la synthèse et

l'accumulation progressive de saccharose dans la tige, commence à partir du 6^{ème} mois et coïncide avec l'abaissement des températures et l'apparition d'un stress hydrique lié à la saison sèche. Un gradient de concentration en saccharose décroissant s'établit du bas vers le haut de la tige. Cette accumulation en sucre est accompagnée d'un ralentissement de la croissance végétative. La date de récolte, qui se situe autour de 12 mois, est fixée en fonction de la valeur du Brix, correspondant à la teneur du jus de la tige en matières sèches solubles (principalement du saccharose et quelques sucres simples) qui est évalué par analyse réfractométrique (Van Dillewijn, 1960 ; Fauconnier, 1991).

I.1.3 Valorisation économique

A la récolte, les tiges sont coupées à leur base et effeuillées. Des techniques de brûlage des feuilles sèches avant la récolte peuvent faciliter celle-ci. Les tiges de canne sont livrées à des sucreries ou des distilleries. La canne traitée en sucrerie donne trois coproduits : le sucre, la bagasse (résidu fibreux) et la mélasse (résidu non cristallisé de la fabrication du sucre) ainsi qu'un sous-produit, l'écume, issu de la clarification du jus de canne utilisé en amendement organique des sols (riche en calcium, azote, phosphore).

Saccharose, mélasse et bagasse sont utilisés respectivement en alimentation humaine, alimentation animale, et combustion et production d'électricité. Ces coproduits peuvent également alimenter des industries associées :

- à partir du saccharose : production de molécules diverses (éthers, esters, polyesters) dans le secteur de la sucro-chimie
- à partir de la mélasse : production de rhums ou d'éthanol industriel par fermentation, production de molécules à haute valeur ajoutée (acides aminés, protéines), production de levures; après distillation de l'alcool sous-production de vinasse (riche en potassium), pouvant servir à la fertilisation des champs ou à la production de cires ou d'acides organiques pour des usages industriels (détergents), pharmaceutiques ou cosmétiques
- à partir de la bagasse : production de pâte à papier, de biomatériaux de construction ou d'ameublement, de textiles. La bagasse est une matière cellulosique intéressante pour

Famille	Sous-famille	Tribu	Genre	Espèce
Poaceae	Chlorideae	Chlorideae	Eleusine	
		Oryzaceae	Oryza	
	Panicoideae	Andropogoneae	Erianthus	
			Miscanthus	
		Saccharum	Saccharum	S. barberi
				S. edule
				S. officinarum
		Paniceae	Sorghum	S. robustum
			Zea	S. sinense
		Pooideae	Pennisetum	S. spontaneum
			Setaria	
		Poae	Festuca	
		Avenae	Lolium	
			Avena	
		Triticeae	Hordeum	
			Secale	
			Triticum	

Figure 4 : Position taxonomique du genre *Saccharum* (adapté de Daniels and Roach, 1987).

la production de biocarburants de seconde génération compte tenu de la biomasse élevée de la canne à sucre. Plusieurs programmes d'amélioration sont aujourd'hui orientés vers la production de « canne à fibre », riche en celluloses et hémicelluloses et pauvre en lignines.

Depuis les années 1970 s'est développée au Brésil (puis en Colombie, au Guyana) une importante filière de bioéthanol de première génération utilisé comme agro-carburant. L'intérêt des agrocarburants en réponse à la raréfaction des ressources pétrolières, est aujourd'hui remis en cause par leur bilan énergétique et environnemental discutable, la compétition avec les cultures à finalités alimentaires, la disponibilité limitée des surfaces agricoles et des ressources en eau (Berndes, 2002). Toutefois, malgré la difficulté d'une évaluation objective de bilan énergétique, la canne à sucre est reconnue comme étant une des plantes les plus intéressantes pour la production de biomasse et de biocarburants, en raison de son rendement photosynthétique élevé.

A partir d'une tonne de canne, on obtient en moyenne 250 à 300 kg de bagasse pour 700 à 800 kg de jus contenant environ 16 % de saccharose (soit 115 kg). La canne, en plus de ses nombreuses applications alimentaires et industrielles, est une culture efficace pour lutter contre l'érosion des sols en zone tropicale pluvieuse, grâce à son système racinaire puissant et à la bonne couverture permanente du sol qu'elle offre.

I.2 Taxonomie et évolution de la canne à sucre

La canne à sucre (*Saccharum* spp.) est une monocotylédone qui appartient à la famille des Poacées qui comprend un grand nombre d'espèces d'intérêt économique. Elle appartient à la tribu des Andropogonées où l'on trouve également le sorgho et le maïs. Au niveau phylogénique, le sorgho est l'espèce cultivée la plus proche de la canne à sucre. Les autres Poacées importantes pour l'alimentation animale et humaine comme le riz, le blé, l'avoine ... appartiennent à d'autres tribus de monocotylédones (**Figure 4**).

I.2.1 Les espèces du genre *Saccharum*

Le genre *Saccharum*, à l'origine de la canne à sucre cultivée, comprend 6 groupes taxonomiques qui ont été définis comme 6 espèces. Ces espèces sont caractérisées par une grande variabilité phénotypique. Toutes possèdent un génome hautement polyploïde de grande taille, des niveaux de ploïdie différents et quelque fois variables au sein d'une même espèce composée de divers cytotypes. Sur la base de critères agro-morphologiques, les botanistes distinguent deux catégories :

- les espèces sauvages, presque dépourvues de sucre
- les espèces domestiquées, productrices de sucre

a. Les espèces sauvages

S. spontaneum est une espèce rustique qui peut s'implanter dans des régions contrastées, tant au niveau de la température et de l'humidité que du type de sol. Cette espèce a une distribution géographique très large qui va de l'Afrique de l'Est au Sud-Est de l'Asie. Son centre d'origine le plus probable se situe en Asie continentale. Sa large répartition géographique témoigne d'une capacité à coloniser différents milieux, qui peut se traduire par un caractère envahissant dans des conditions favorables. *S. spontaneum* est une plante annuelle, pérenne ou semi-pérenne, qui peut se reproduire par voie sexuée ou végétative et produit des rhizomes. Elle est caractérisée par des tiges fines, pauvres en sucre et riches en fibres, et un tallage important, ainsi qu'une grande diversité génétique et phénotypique, et une tolérance à différents stress biotiques et abiotiques. Sa taille varie en fonction des écotypes et elle peut atteindre jusqu'à 5 m de haut. Le nombre de chromosomes varie entre $2n=40$ et $2n=128$, avec un nombre chromosomique de base $x=8$ et un niveau de ploïdie variant de 5 à 16 en fonction des cytotypes (D'hont *et al.*, 1996 ; D'Hont *et al.*, 1998). Les cytotypes les plus fréquents sont des multiples de 8 (Panje et Babu, 1960). Le génome haploïde de l'espèce est estimé à environ 760 Mpb (Butterfield *et al.*, 2001).

S. robustum, l'autre espèce sauvage du genre *Saccharum*, ne produit pas de rhizome. Son inflorescence est plus grande que celle de *S. spontaneum*, ses épillets plus petits, et elle a une biomasse importante avec des tiges longues, épaisses et très ligneuses. *S. robustum* est

une espèce endémique de Nouvelle-Guinée et de Mélanésie. Elle a un nombre chromosomique de base de $x=10$ (D'Hont *et al.*, 1998). Deux cytotypes euploïdes ont été identifiés : $2n=60$ ou $2n=80$, ainsi que des cytotypes aneuploïdes variant de $2n=63$ à 205.

b. Les espèces domestiquées

S. officinarum, l'espèce sucrière souvent appelée « espèce noble », est issue de la domestication de *S. robustum* et a été la principale espèce cultivée pendant des siècles. Elle est encore utilisée dans des agricultures traditionnelles, en particulier en Mélanésie. Le centre de diversité de cette espèce se situe en Nouvelle-Guinée (Daniels and Roach, 1987). Elle possède de grosses tiges au port érigé, remarquables par leur richesse en sucre, ainsi que leur faible teneur en fibres et en amidon (Bremer, 1961 ; Roach 1986). Elle a des caractéristiques agronomiques variables suivant les clones (Daniels and Roach, 1987), et une diversité génétique appréciable (Jannoo *et al.*, 1999b). C'est une espèce euploïde à $2n=80$ chromosomes (Price and Daniels, 1968), octoploïde, avec un nombre chromosomique de base $x=10$, et un génome haploïde d'environ 926 Mpb (Butterfield *et al.*, 2001)

Les espèces *S. barberi* et *S. sinense* sont deux petits groupes horticulturaux respectivement cultivées en Inde du Nord (l'est de l'actuel Pakistan) et en Chine du Sud. Les clones appartenant à ces deux groupes sont des hybrides naturels de premières générations entre des clones appartenant à *S. spontaneum* et des clones appartenant à *S. officinarum* (D'Hont *et al.*, 2002). Ces clones diffèrent des *S. officinarum* par leurs caractères floraux, leur teneur en fibre élevée, une moindre teneur en sucre et une rusticité accrue (meilleure adaptation aux environnements subtropicaux) due à leur nature hybride. Le nombre chromosomique de base varie entre $2n=84$ et $2n=124$, et les méioses sont irrégulières.

S. edule est une espèce stérile constituée d'un petit groupe de clones, très semblables aux clones *S. robustum*. Elle est cultivée en Nouvelle-Guinée et dans les îles Fidji pour ses inflorescences avortées comestibles. Son nombre chromosomique de base varie entre $2n=60$, 70 ou 80 (Roach, 1972 ; Daniels and Roach, 1987).

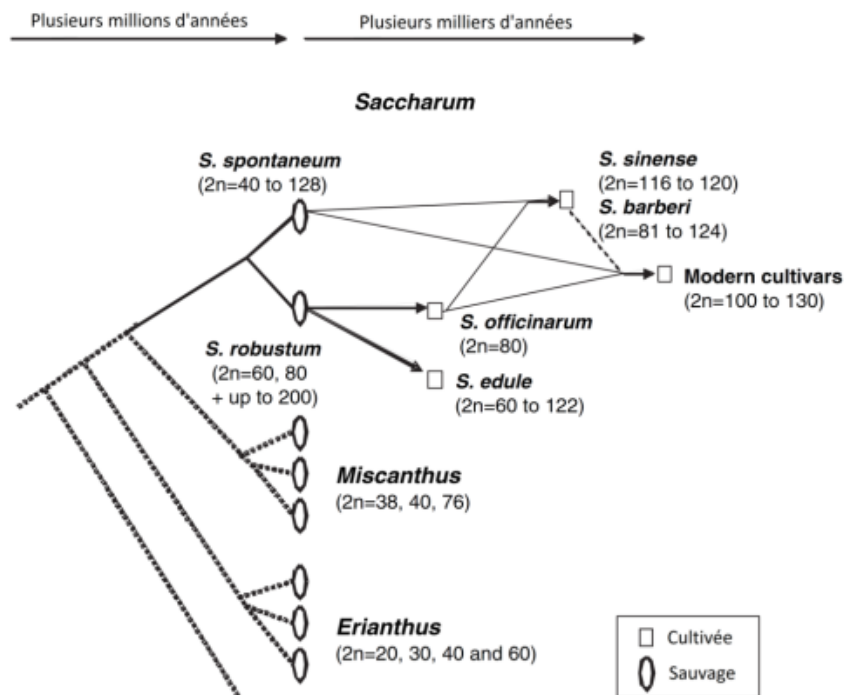


Figure 5 : Evolution et domestication de la canne à sucre : scénario proposé par D'Hont *et al.* (2008) et adapté de Grivet *et al.* (2004, 2006), compatible avec les données moléculaires récentes.

I.2.2 Evolution dans le « complexe *Saccharum* »

L'évolution au sein du genre *Saccharum* ainsi que les liens phylogénétiques entre le genre *Saccharum* et quelques genres apparentés, regroupés dans le concept de « complexe *Saccharum* » (Mukherjee, 1957 ; Daniels *et al.*, 1975), est source d'interrogations depuis longtemps. En 1987, Daniels and Roach ont publié une synthèse détaillée de scénarios d'évolution et de domestication, sur la base de connaissances botaniques, génétiques, cytologiques ou historiques. Récemment le Cirad a revisité les scénarios émis jusqu'à présent, en réunissant une synthèse de données moléculaires provenant d'études de diversité et fondées sur des marqueurs nucléaires, cytoplasmiques ou biochimiques (Grivet *et al.*, 2004 ; 2006). Ce nouvel éclairage apporté par la génétique moléculaire a permis de clarifier certains débats. Le scénario d'évolution de la canne à sucre le plus réaliste proposé par D'Hont *et al.* (2008), qui est reproduit à la **Figure 5**, montre que :

- Les genres *Erianthus* et *Miscanthus* correspondent à des phylums distincts de celui de *Saccharum* qui ont divergé depuis une longue période évolutive. L'hypothèse d'une contribution récente de ces 2 genres apparentés à l'émergence de la canne à sucre est infirmée ;
- Au sein de *Saccharum*, la divergence initiale distingue *S. robustum* de *S. spontaneum*.
- *S. robustum* est à l'origine de *S. edule* et de *S. officinarum*
- Les cultivars modernes sont issus de croisements entre *S. spontaneum* et *S. officinarum* soit directement, soit indirectement via *S. sinense* ou *S. barberi*.

I.3 Amélioration génétique de la canne à sucre

Jusqu'à la découverte de la fertilité de la canne et avant les premières hybridations interspécifiques, la conservation par multiplication végétative de variants naturels, ou domestication, a constitué la seule méthode de sélection utilisée pour l'amélioration de la canne à sucre. Initialement, la plante était essentiellement cultivée pour son jus, les tiges étant mâchées pour en extraire le jus. La préférence des tribus primitives de Nouvelle-Guinée pour des clones plus sucrés, moins fibreux et aux tiges colorées a conduit à la conservation et à la multiplication de clones présentant ces caractères (Stevenson, 1965).

Jusqu'au milieu du XIX^{ème} siècle, la culture de la canne reposait sur quelques clones appartenant aux espèces domestiquées sucrières, *S. officinarum* et *S. sinense-barberi*. Ces clones étaient propagés et conservés par multiplication végétative, car réputés pour se reproduire uniquement par clonage à cause de leur stérilité mâle. La fertilité de la canne a été découverte à Barbade en 1858. Les premières variétés hybrides sont mentionnées à Barbade, Martinique, Java, Réunion et Maurice, mais ont suscité peu d'intérêt car elles étaient pauvres en sucre. Cependant, pendant la deuxième moitié du XIX^{ème}, l'émergence de maladies, notamment la mosaïque, liées à la culture intensive d'un trop petit nombre de clones sensibles appartenant aux espèces domestiquées (Maurice, Réunion, Brésil, Caraïbes) a poussé les planteurs à rechercher et produire de nouvelles variétés (Stevenson, 1965).

Les premiers croisements interspécifiques, réalisés entre des porte-graines « nobles » *S. officinarum*, et des clones *S. spontaneum* sauvages utilisés comme parent mâle, suivis d'un ou deux rétrocroisements sur des parents femelles *S. officinarum* ont donné des premiers hybrides vigoureux et résistants aux maladies. Dans ces schémas d'introggression appelés « nobilisation » chez la canne à sucre, le retour à des clones de bonne valeur sucrière s'est effectué d'autant plus rapidement que les parents *S. officinarum* transmettaient des gamètes diploïdes ($2n$) lors du croisement interspécifique et du premier rétrocroisement (Bhat and Gill, 1985). Les cultivars modernes sont tous des produits de croisements successifs effectués à partir d'un nombre limité de fondateurs interspécifiques de premières générations obtenus au début du XX^{ème} siècle. L'étude de la généalogie des 107 cultivars les plus répandus dans le monde en 1965 dont dérivent la plupart des cultivars contemporains, a montré l'implication d'un nombre limité de clones ancestraux fondateurs : 19 clones *S. officinarum*, quelques clones *S. spontaneum*, 1 clone *S. barberi* et 1 clone *S. robustum*, parmi lesquels 4 clones *S. officinarum* et 2 clones *S. spontaneum* interviennent en fréquence très forte (Arcenaux, 1965). Cette base génétique étroite, combinée à un faible nombre de méioses séparant ces fondateurs des clones modernes, est à l'origine du déséquilibre de liaison relativement important existant parmi le matériel contemporain (Jannoo *et al.*, 1999b ; Raboin *et al.*, 2008).

Le faible nombre de clones *S. spontaneum* impliqués dans la généalogie des cultivars modernes et la diversité génétique importante révélée chez cette espèce (Lu *et al.*, 1994 ; Pan *et al.*, 2004) permettent d'affirmer qu'une grande part de la diversité génétique existant dans *Saccharum* n'a pas encore été exploitée jusqu'à présent. Par ailleurs la possibilité de croisements inter-génériques avec certains genres apparentés à *Saccharum* (*Erianthus*,



Figure 6 : Floraison et hybridation : a) champ de canne en période de floraison, b) fleurs de canne en chambre de croisement, c) détail d'une fleur avec stigmates déhiscent d) fleur porte-graine femelle avec stigmates indéhiscent [Photos a,d,e) D. Roques, CIRAD ; b) J.Y. Hoarau, CIRAD].

Miscanthus...) peut offrir une source supplémentaire de diversité potentiellement utile à l'amélioration de la canne à sucre (Daniels and Roach, 1987 ; Cai *et al.*, 2005).

I.3.1 Floraison et hybridation

L'aptitude à la floraison est variable selon les variétés. Un nombre important de variétés ne fleurissent jamais même dans des conditions écologiques favorables, et parmi celles qui fleurissent, il existe une grande variabilité des taux de floraison. De plus, la floraison provoquant un arrêt de la croissance végétative et une diminution de la teneur en sucre de la tige, donc du rendement (Berding and Hurney, 2005), ce caractère a été largement contre-sélectionné, à l'exception d'un taux de floraison résiduel nécessaire au maintien des possibilités de croisements.

L'initiation de la floraison est sous contrôle photopériodique. Elle est induite par les jours courts et divers facteurs environnementaux tels que la baisse des températures et certains stress abiotiques. La date de floraison dépend de la latitude et varie en fonction des espèces et variétés. Il existe une différence d'environ 3 mois entre les clones les plus précoces et les plus tardifs. La floraison d'une variété donnée s'échelonne sur une période stable d'une année sur l'autre qui dure environ 15 jours. Plusieurs techniques permettent de synchroniser les périodes de floraison de différents clones (éclairage artificiel ou plantation en altitude d'un clone tardif...), et d'augmenter les taux de floraison (stress hydrique ou azoté...) afin de réaliser des croisements inter-variétaux recherchés.

L'inflorescence de la canne à sucre est organisée en panicule regroupant une multitude d'épillets hermaphrodites, typique des Poacées, comprenant trois anthères à déhiscence mécanique, deux stigmates plumeux surmontant un ovaire uniloculaire et unilovulé (**Figure 6**). La pollinisation est anémophile. La fertilité sporophytique, liée à la déhiscence des anthères, et la fertilité pollinique, liée à la capacité de germination du grain de pollen, sont très variables d'un clone à un autre. On considère un génotype comme femelle (porte-graine) lorsque sa fertilité pollinique est inférieure à 10 %. Au-delà de 20 % de fertilité pollinique, le génotype est utilisable comme parent pollinisateur. La canne est préférentiellement allogame mais peut tolérer l'autofécondation pendant une ou deux générations. Compte tenu de

l'hétérozygotie élevée des géniteurs, et de la forte dépression de consanguinité observée lors d'autofécondations successives, la production de lignées pures est impossible.

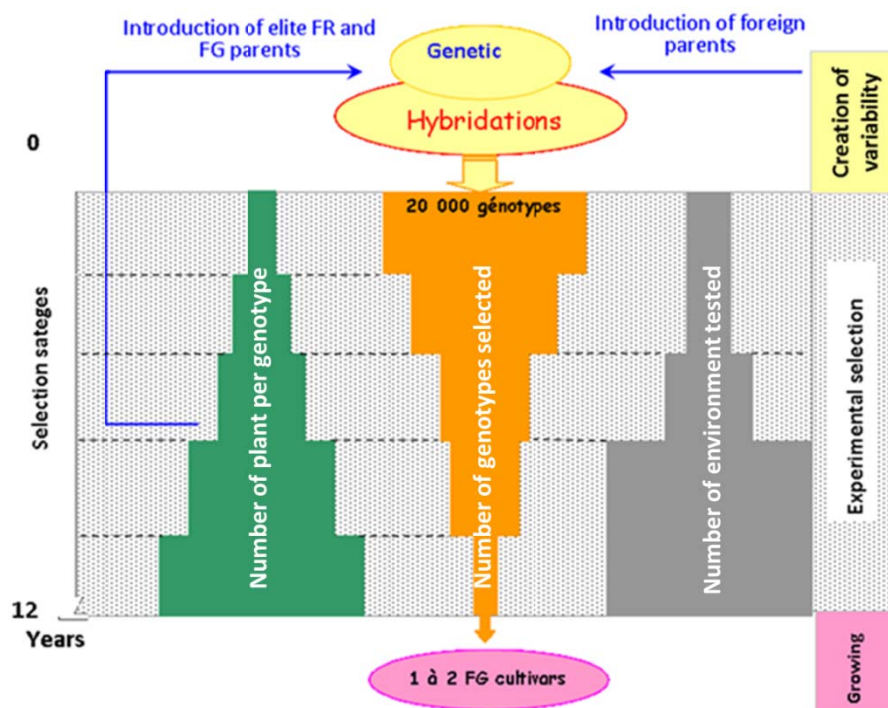
I.3.2 Critères de sélection

La création variétale chez la canne à sucre a pour but de produire des génotypes productifs, tolérants aux stress abiotiques et résistants aux maladies. La sélection prend en compte des critères agronomiques, de qualité technologique des cannes et de résistance aux maladies.

Le critère agronomique principal est la production de biomasse aérienne, essentiellement déterminé par le rendement de canne à sucre usinable (tonnage par unité de surface) et ses composantes agro-morphologiques (diamètre, tallage, hauteur, absence de floraison). Cette notion de rendement s'analyse aussi dans le temps (pérennité du tonnage en repousses après les coupes successives) et dans l'espace (adaptation à diverses écologies de culture). Les autres caractères agronomiques examinés concernent des aspects liés à la limitation du coût et à la facilité du management phytotechnique et de la récolte mécanique (port érigé des tiges, rapidité de croissance, importance de la couverture foliaire du sol).

Les critères de qualité technologique de la canne varient selon les filières économiques considérées. Pour les filières sucre et éthanol de première génération, on privilégiera une teneur importante en sucre du jus et une faible teneur en fibres de la canne, pour faciliter le processus d'extraction du jus et obtenir une richesse optimale de ce dernier. Dans le cadre de la production de biomasse destinée à la production d'électricité, on privilégiera une teneur en fibres ligno-cellulosiques élevée associée à un pouvoir calorifique important, et des teneurs en sucres et en eau moindre pour faciliter la combustion. Pour des filières cherchant à valoriser la cellulose de la bagasse, l'idéotype recherché contiendrait une faible teneur relative en lignine pour faciliter l'extraction de la cellulose (pâte à papier) et/ou la saccharification de la cellulose (éthanol de seconde génération).

La résistance aux bio-agresseurs constitue un critère de sélection essentiel, notamment s'agissant des maladies qui affectent la tige de façon systémique. C'est le cas de toutes les maladies virales et bactériennes, et de quelques maladies fongiques. Les éventuelles



Principe du schéma de sélection des variétés FR/FG

Figure 7 : Représentation schématique du processus de sélection clonale de la canne à sucre pratiqué en Guadeloupe par étapes expérimentales successives de plus en plus sélectives à partir d'une production annuelle de 20 000 génotypes (Oriol and Domaingue, 2008).

dispositions préventives à la plantation (traitement thermique des boutures, régénération de boutures saines par culture *in vitro*) ne sont pas efficaces longtemps s'il s'agit de matériel sensible. Compte tenu de la nécessité d'avoir des rendements pérennes sur plusieurs cycles de culture, la sélection de cultivars à des niveaux de résistance élevés est considérée comme la seule méthode de lutte efficace contre la propagation des maladies, qui soit respectueuse de l'environnement et acceptable d'un point de vue économique.

I.3.3 Principe des schémas de sélection

La sélection variétale de la canne est basée sur un principe de sélection récurrente, qui vise à progressivement améliorer des populations sur un ensemble de caractères, par des cycles successifs d'hybridation suivis de sélection. Le choix des combinaisons sexuées est déterminé par la valeur propre des clones, leur aptitude particulière ou générale à la combinaison, ainsi que par les contraintes de floraison (fertilité pollinique et apparition de panicules synchrones).

Chaque année un très grand nombre de clones issus de centaines de croisements inter-variétaux sont créés puis évalués sur plusieurs années dans un schéma de sélection familiale puis clonale, comparable à celui pratiqué chez d'autres plantes à multiplication végétative (pomme de terre, bananier, tulipe ...). La première étape de la sélection consiste à ne conserver que les seules graines viables. Les étapes suivantes consistent à sélectionner de manière progressive à travers quatre à cinq étapes de plus en plus sélectives les clones les plus prometteurs, afin de les tester à plus grande échelle et sur une plus longue durée (évaluation de la tenue en repousse). La sélection porte d'abord sur les caractères les plus héréditaires (teneur en Brix, taux de floraison, maladies cryptogamiques foliaires) puis sur ceux qui le sont moins (tonnage) ou qui sont laborieux d'évaluation (diagnostics pathologiques en laboratoire).

Au cours des étapes du schéma de sélection, il y a une diminution progressive du nombre de clones testés parallèlement à une augmentation de la surface, du nombre de parcelles observées pour chaque clone et du nombre d'environnements écologiques expérimentés (**Figure 7**). Un cycle de sélection dure environ 10 à 12 ans, et aboutit à la diffusion des meilleures variétés. Le principe de récurrence est lié à la réutilisation des descendants élites comme géniteurs dans le programme d'hybridation. Ces élites commencent

à être identifiables à mi-parcours du schéma (environ 6 ans) (J.Y. Hoarau, communication personnelle).

I.4 Les maladies de la canne à sucre

I.4.1 Panorama général

a. Importance de la sélection contre les maladies

La canne à sucre, est une plante exploitée de manière semi-pérenne. La durée des cycles est de 12 mois en moyenne, et quatre à huit cycles de culture successifs peuvent être réalisés pour une plantation. La durée et le nombre de cycles de culture varient selon les terroirs, les modes de culture pratiqués et l'aptitude à la repousse des variétés utilisées. Le coût important de la plantation initiale est d'autant mieux amorti que la pérennité des champs est longue. Cette pérennité dépend notamment des risques d'apparition d'épidémies, qui peuvent être limités par une stratégie d'association de parcelles monovariétales distinctes, composées de géotypes différents possédant des niveaux de résistance élevés et idéalement d'origines diverses. Les producteurs recherchent toujours des clones possédant des niveaux de résistance relativement élevés aux maladies les plus dommageables. Cette exigence est d'autant plus importante en l'absence de méthodes de traitements curatifs viables et acceptables sur les plans économiques, environnementaux et techniques.

b. Abondance de maladies systémiques potentiellement dommageables.

Au total on dénombre dans le monde plus de 120 maladies sur la canne à sucre, parmi lesquelles une soixantaine peut causer des dommages conséquents dans certaines conditions environnementales favorables (Rott *et al.*, 2000). Les maladies les plus dommageables pour les rendements sont celles d'origine tellurique qui affaiblissent le système racinaire, ainsi que celles qui infectent de façon systémique tous les organes de la plante, et qui de surcroît, se propagent par l'intermédiaire de boutures et persistent d'un cycle de culture à l'autre. Parmi les maladies fongiques qui touchent uniquement les feuilles, les plus dommageables sont celles atteignant les feuilles jeunes (les plus actives sur le plan photosynthétique). Le nombre et l'importance économique des maladies est variable selon les pays et l'environnement de culture. Les efforts de lutte génétique contre les maladies, conçus dans les programmes d'amélioration, doivent être adaptés à chaque contexte écologique.

Tableau 2 : Maladies de la canne recensées en Guadeloupe (J.H. Daugrois, communication personnelle).

Nom	Agent causal	Vecteur	Hôte alternatif
Maladies virales			
Stries chlorotiques	inconnu		Nombreuses graminées
Mosaïque*	SCMV SRMV, Potyvirus	<i>Aphids</i>	Nombreuses graminées
Sugarcane Bacilliform Virus	SCBV <i>Badnavirus</i>	<i>Saccharicoccus sacchari</i>	<i>Sorghum halepense</i> , <i>Sorghum vulgare</i> , <i>Brachiaria sp.</i> , <i>Panicum maximum</i> , <i>Rottboellia exaltata</i> , <i>Oryza sativa</i> , <i>Musa 'AAA'</i>
Feuille jaune	SCYLV <i>Polerovirus</i>	<i>Melanaphis sacchari</i>	-
Maladies bactériennes			
Échaudure des feuilles	<i>Xanthomonas albilineans</i>		<i>Zea maize</i> , nombreuses graminées
Striure marbrée (Mottled stripe)	<i>Herbaspirillum rubrisubalbicans</i>		<i>Sorghum spp.</i>
Rabougrissement des repousses	<i>Leifsonia xyli</i> subsp. <i>xyli</i>		-
Striure rouge	<i>Acidovorax avenae</i> subsp. <i>avenae</i>		Graminées cultivées (<i>Zea Maize</i> , <i>Sorghum bicolor</i> , millet)
Pourriture en fuseau (Spindle rot)	related to <i>Acidovorax avenae</i> subsp. <i>avenae</i>		-
Maladies fongiques			
Raies brunes	<i>Bipolaris stenopila</i>		-
Rouille commune	<i>Puccinia melanocephala</i>		<i>Erianthus</i> sp., <i>Narenga porphyrocoma</i> , <i>Bambusa sp.</i>
Pourriture de tige et bouture	<i>Giberella fuikuroi</i> = <i>Fusarium moniliforme</i>		Nombreuses monocotyledones et dicotyledones
Crise de pourriture des tiges, racines et gaines (Marasmius stem, root and sheat rot)	<i>Marasmius sacchari</i> / <i>Marasmius stenospilus</i>		-
Maladie de l'ananas	<i>Ceratocystis paradoxa</i>		Nombreux (<i>Musa spp</i> , <i>Theobroma cacao</i> , <i>Cocos nucifera</i> , <i>Zea maize</i> , <i>Elaeis guineensis</i> , <i>Phoenix canariensis</i> , <i>Ananas comosus</i>)
Pokkah boeng	<i>Giberella fuikuroi</i> / <i>Giberella subglutinans</i>		Nombreuses monocotyledones et dicotyledones
Morve rouge	<i>Glomerella tucumanensis</i>		Nombreuses graminées sauvages, <i>Sorghum spp.</i> , <i>Miscanthus spp.</i>
Tache rouge de la gaine foliaire (Red spot on the leaf sheath)	<i>Mycovellosiella vaginiae</i>		-
Maladie de l'écorce (Rind disease)	<i>Phaeocytostroma sacchari</i>		-
Tache annulaire	<i>Leptosphaeria sacchari</i>		-
Charbon	<i>Ustilago scitaminea</i>		<i>Miscanthus spp.</i> , <i>Zea maize</i> , <i>Sorghum bicolor</i> , graminées sauvages (<i>Chloris barbata</i> , <i>cynodon dactylon</i> , <i>Imperata cylindrica</i> , <i>Cyperus rotundus</i>)
Flétrissement	<i>Pathogènes multiples</i>		-
Tache jaune	<i>Mycovellosiella koepkei</i>		<i>Miscanthus sinensis</i> , <i>Miscanthus floridulus</i> , graminées sauvages
Fumagine	<i>Capnodium sp.</i>		-

* n'a pas été observée depuis plus de 20 ans

Parmi les maladies les plus importantes (Fauconnier, 1991 ; Rott *et al.*, 2000), on peut citer dans le domaine des maladies fongiques, le charbon (*Ustilago scitaminea*) et la rouille brune (*Puccinia melanocephala*). La maladie du charbon, qui existe dans tous les pays sucriers, est caractérisée par l'apparition de fouets qui provoquent d'importants dégâts. Elle est disséminée par des spores pouvant se propager sur de très longues distances. La rouille brune est une maladie très pénalisante chez les clones sensibles cultivés dans les zones aux températures relativement fraîches. Deux gènes majeurs de résistance à la rouille brune ont été identifiés dans des cultivars de canne (Daugrois *et al.*, 1996 ; Raboin *et al.*, 2006). Parmi les maladies bactériennes, le rabougrissement des repousses (*Leifsonia xyli* susp *xyli*), la gommose (*Xanthomonas axonopodis* pv *vasculorum*) et l'échaudure des feuilles (*Xanthomonas albilineans*) sont des maladies redoutables en raison d'une symptomatologie discrète (rabougrissement), qui ne s'exprime pas nécessairement juste après la contamination systémique des tiges (gommose, échaudure des feuilles). Dans le domaine des maladies virales, deux maladies transmises par plusieurs espèces de vecteurs affectent de très nombreux pays et sont susceptibles de causer des dégâts significatifs : la mosaïque (*Sugarcane Mosaic Virus*, SCMV) et la feuille jaune (*Sugarcane Yellow Leaf Virus*, SCYLV).

Vingt-cinq maladies de la canne ont été référencées en Guadeloupe, parmi lesquelles quatre font l'objet de travaux en sélection, en génétique et en épidémiologie (charbon, échaudure des feuilles, feuille jaune, rabougrissement des repousses) (**Tableau 2**). Des études d'associations concernant la résistance à l'échaudure des feuilles et à la feuille jaune sont entreprises dans le présent travail de thèse. Les principales caractéristiques de ces deux maladies et de leur pathosystème sont détaillées dans les paragraphes suivants.

I.4.2 L'échaudure des feuilles

a. Impact économique, agent causal et gamme d'hôtes

L'échaudure des feuilles de la canne à sucre a été recensée dans 66 pays producteurs de canne (Rott and Davis, 2000) et peut causer des baisses de rendement importantes ou une diminution de la teneur en sucre chez des cultivars sensibles (Rott *et al.*, 1995, Hoy, 1994, Holder *et al.*, 2000). Cette maladie est provoquée par une bactérie vasculaire : *Xanthomonas albilineans*. Il s'agit d'une bactérie Gram négatif, aérobie, non sporulante, en forme de

bâtonnet et possédant un unique flagelle polaire comme toutes les bactéries pathogènes du genre *Xanthomonas*. Isolée en culture pure, elle peut être identifiée selon les méthodes habituellement utilisées en phytobactériologie (Rott *et al.*, 1988 ; Davis *et al.*, 1997). Elle peut aussi être détectée par des méthodes sérologiques (Comstock and Irey, 1992 ; Davis *et al.*, 1997 ; Tsai *et al.*, 1992) ou par test PCR (Pan *et al.*, 1997 ; Lopes *et al.*, 2001a). En règle générale, *X. albilineans* colonise essentiellement les vaisseaux du xylème, et sa présence ainsi que les polysaccharides qu'elle produit (Solas *et al.*, 2003), qui obstruent les vaisseaux les plus fins, perturbent l'alimentation hydrique de la plante (Rott *et al.*, soumis).

Le génome de la bactérie a été récemment séquencé et annoté à partir de l'étude d'une souche provenant de Guadeloupe (Pieretti *et al.*, 2009). Ce génome est constitué d'un seul chromosome circulaire de plus de 3 millions de paires de bases et de trois plasmides de quelques dizaines de paires de bases. *X. albilineans* est proche d'un point de vue phylogénétique de *Xylella fastidiosa* (Simpson *et al.*, 2000 ; Pieretti *et al.*, 2009), et colonise également uniquement le xylème. *X. albilineans* possède la particularité d'être dépourvue de gènes *hrp* caractéristiques des bactéries à Gram négatif (Büttner and Bonas, 2010). Ces gènes sont impliqués dans la pathogénicité de nombreuses bactéries et dans la réponse hypersensible (HR) des plantes. L'identification des facteurs de pathogénicité spécifiques de la bactérie progresse actuellement grâce à l'analyse d'une collection de mutants d'insertions de *X. albilineans*, dont certains sont affectés dans leur capacité à se propager *in planta* et à produire des symptômes foliaires (Rott *et al.*, 2011).

X. albilineans est une bactérie inféodée à la canne à sucre mais, selon Birch (2001), elle peut aussi infecter naturellement le maïs ainsi que diverses graminées adventives situées à proximité de cannes malades. Le sorgho serait susceptible d'être inoculé artificiellement.

b. Expression des symptômes

Dans la symptomatologie, on peut distinguer des symptômes foliaires consécutifs à des infections systémiques de la tige, et sous certains climats humides favorables, des symptômes foliaires provoqués par une contamination des feuilles issue d'une propagation aérienne de la bactérie. Les symptômes foliaires résultant d'une infection systémique d'origine interne, sont toujours erratiques, souvent discrets et fugaces ; alors que les symptômes foliaires provenant d'une contamination aérienne externe pourraient s'exprimer de façon plus homogène et être observables sur une plus longue durée.

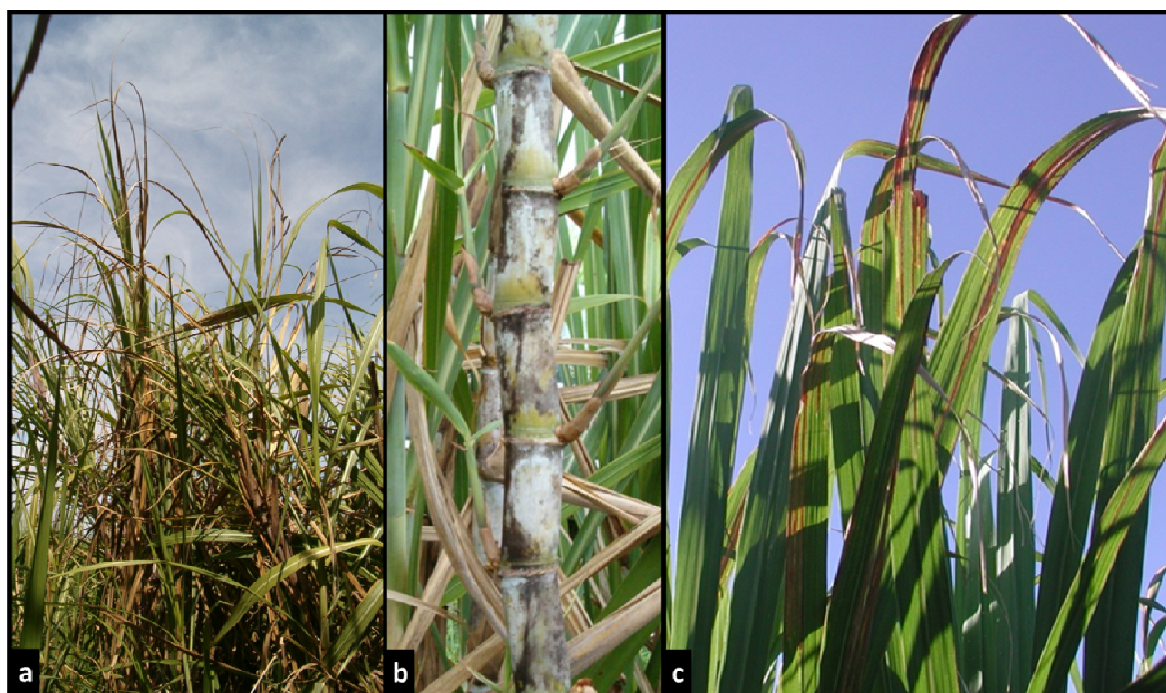


Figure 8 : Echaudure des feuilles de canne à sucre. a) Forme aigüe ; b) Forme systémique ; c) Symptômes caractéristiques d'une contamination par voie aérienne (Photos : J.H. Daugrois, CIRAD).

- **Symptômes foliaires résultant d'une infection d'origine interne**

On peut distinguer une **forme aigüe** et une **forme chronique**.

La **forme aigüe** se manifeste par des symptômes semblables à ceux que produirait un stress hydrique intense : les tiges infectées flétrissent brutalement, le plus souvent sans symptômes avant-coureurs. Cette forme aigüe de la maladie, qui peut conduire à la mort de la plante, n'apparaît que chez les variétés très sensibles et sur des tiges matures, souvent après une période de déficit hydrique (Rott and Davis, 2000) (**Figure 8a**).

La **forme chronique** se manifeste classiquement par l'apparition sur les feuilles d'une ou de plusieurs fines lignes blanches bordées de zones jaunes, qui partent de la nervure centrale et se prolongent sur le limbe. Ces lignes correspondent à la colonisation de faisceaux vasculaires par la bactérie. Elles peuvent s'élargir vers l'extrémité des feuilles et se terminer par des chloroses ou des nécroses à un stade avancé de la maladie. Le développement de tiges et racines à partir des bourgeons axillaires peut également être observé sur les tiges matures. L'expression de cette forme dite chronique de la maladie est erratique : d'une part, l'apparition des symptômes peut être précédée d'une phase de latence de plusieurs mois, pendant laquelle les cannes infectées de manière systémique peuvent ne montrer aucun symptôme ; d'autre part, les symptômes ne sont généralement pas permanents (les symptômes portés par des feuilles particulières disparaissent au moment de la sénescence de celles-ci), et on observe souvent des « phases d'éclipse » selon le rythme de croissance de la plante et les fluctuations climatiques. Dans une expérimentation tous les génotypes sensibles ne sont donc pas identifiables à une même et unique date d'observation (Rott and Davis, 2000) (**Figure 8b**).

- **Symptômes résultant d'une contamination aérienne**

Depuis une vingtaine d'années, des épidémies ont été signalées dans plusieurs régions du monde où la maladie n'existait pas (Ovalle *et al.*, 1995 ; Chen *et al.*, 1993 ; Isakeit and Irvine, 1995) ou ré-apparaissait après avoir été efficacement contrôlée par la lutte génétique (Comstock and Shine, 1992 ; Rott *et al.*, 1994 ; Autrey *et al.*, 1995 ; Diaz *et al.*, 2001). Jusqu'alors, la contamination à l'échaudure des feuilles était exclusivement attribuée à la propagation de boutures infectées de cannes sensibles, et aux instruments de coupe. A la faveur de ces nouvelles épidémies, un nouveau mode de propagation de la bactérie par voie

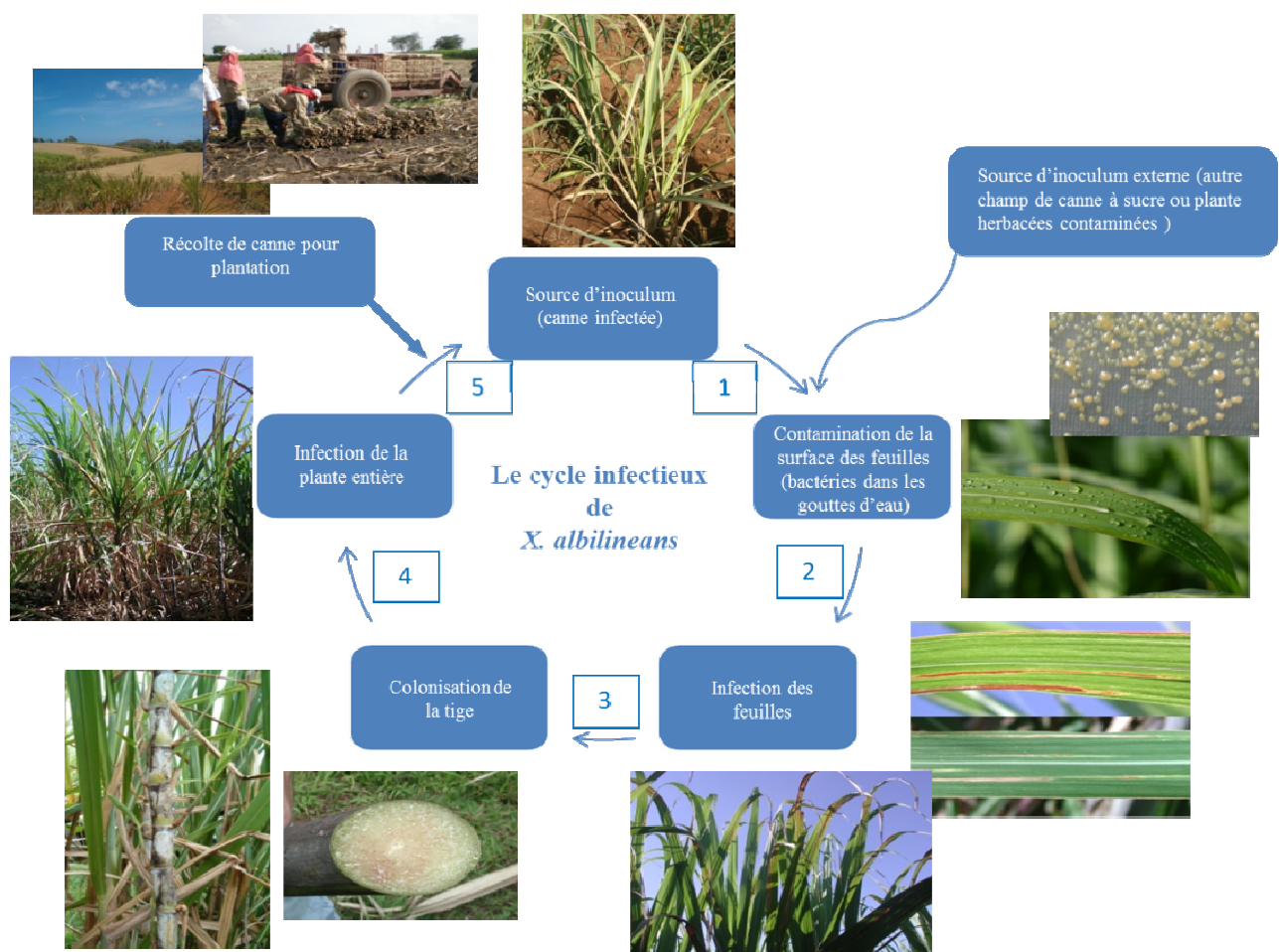


Figure 9: Cycle infectieux de *X. albilineans* (d'après Daugrois *et al.*, in press).

aérienne fut découvert (Klett and Rott, 1994 ; Autrey *et al.*, 1995 ; Comstock, 2001 ; Daugrois *et al.*, 2003). Des populations bactériennes existant à la surface des feuilles en saison humide peuvent se transmettre par simple contact entre feuilles, ou par la pluie ou les gouttes de rosée. Les symptômes foliaires observés sont caractérisés par de fines nécroses linéaires, souvent bordées de zones jaunâtres, qui se développent sur le limbe ou sur ses bords à partir d'une pénétration initiale de la bactérie par des stomates ou des blessures (**Figure 8c**). Les lésions nécrotiques qui atteignent la nervure centrale des feuilles peuvent conduire à une primo infection de la tige si cette dernière est saine ou à une surinfection dans le cas contraire. Le nombre et la taille des lésions nécrotiques varient selon les cultivars, les variétés les plus sensibles présentant les symptômes les plus étendus (Comstock, 2001).

Une synthèse récente de nouvelles connaissances acquises en matière d'épidémiologie a permis de reconsidérer une description plus réaliste du cycle infectieux de la bactérie, qui tient compte de l'importance de la contamination par voie aérienne (**Figure 9**) (Daugrois *et al.*, in press). Champoiseau *et al.* (2009) ont montré que chez le cultivar B69566, en présence d'inoculum naturel, il existe une corrélation entre l'intensité des nécroses foliaires et la colonisation épiphyte de la plante par *X. albilineans*. Dans des conditions de pluies suffisamment abondantes, le taux de tiges contaminées par *X. albilineans* après colonisation de la phyllosphère est principalement lié à la densité de population bactérienne sur les feuilles. Cette dernière est favorisée par la présence de souches bactériennes particulières, adaptées à la survie par voie épiphyte (Davis *et al.*, 1997). Si la stabilité et la durabilité des symptômes nécrotiques foliaires pouvaient être démontrées, ces derniers pourraient constituer un indicateur de choix pour apprécier le niveau de sensibilité d'une variété à la contamination aérienne de la plante par *X. albilineans* (J.H. Daugrois, communication personnelle).

c. Diversité génétique du pathogène

La diversité génétique de *X. albilineans* a été étudiée par diverses techniques de marquage aléatoires (Davis *et al.*, 1997 ; Lopes *et al.*, 2001a ; Champoiseau *et al.*, 2006b ; Shaik *et al.*, 2009) ainsi qu'au niveau de gènes impliqués dans la biosynthèse d'une phytotoxine, l'albicidine (Champoiseau *et al.*, 2006a et 2006b ; Renier *et al.*, 2007). En combinant ces diverses approches de marquage Champoiseau *et al.* (2006b) ont montré l'existence en Guadeloupe de souches génétiques peu différenciables génétiquement les unes

des autres, en dépit de l'existence d'une large diversité du pouvoir pathogène entre ces souches.

Davis *et al.* (1997), dans une étude de diversité portant sur un large échantillon mondial, ont démontré au moyen de profils de restriction sur le génome entier que toutes les épidémies survenues récemment dans de nombreuses régions du monde [Floride, Guadeloupe, Louisiane, Cuba, Texas, Taiwan, Afrique du Sud] ont été provoquées par des souches bactériennes appartenant à un même et unique groupe génétique (le groupe B). Ce groupe correspond vraisemblablement à des populations bactériennes possédant des capacités de survie épiphyte et de transmission aérienne supérieures à celles des autres groupes (Davis *et al.*, 1997). La découverte de l'appartenance à ce même groupe génétique B des souches responsables de ces épidémies récentes a été confirmée par une deuxième étude utilisant des marqueurs RFLP (Champoiseau *et al.*, 2006a). Par ailleurs cette deuxième étude a révélé que toutes les souches bactériennes dernièrement échantillonnées en Guadeloupe font partie de ce groupe B.

d. Sources de résistance de l'hôte et nature du déterminisme

Parmi les hybrides interspécifiques modernes on peut facilement sélectionner des clones très résistants utilisables comme géniteurs, soit sur une base uniquement symptomatologique après inoculation naturelle (Egan, 1972 ; Bechet *et al.*, 1992) soit par des mesures de symptômes et de colonisation des tiges après une inoculation artificielle (Rott *et al.*, 1994 ; Lopes *et al.*, 2001b).

Parmi les espèces ancestrales, des observations symptomatologiques (Egan, 1972) et de densités de colonisation (Daugrois *et al.*, 1995 ; Daugrois *et al.*, 1998) concordantes montrent que la plupart des clones sensibles appartiennent aux espèces *S. robustum* et *S. officinarum* et la plupart des clones résistants se trouvent chez *S. spontaneum*.

Les sources de résistance présentes dans le germplasm (*Saccharum* spp.) sont en définitive nombreuses. Toutefois il existe peu d'information dans la littérature sur la nature des déterminismes sous-jacents et leur transmissibilité en croisement. Davantage d'informations seraient très utiles à l'orientation rationnelle des efforts d'amélioration.

Concernant la nature des déterminismes, la seule étude disponible concerne l'analyse génétique de la résistance du clone 'R570' abordée au moyen d'une étude QTL (Offmann, 2000). Ce clone qui peut montrer des symptômes épisodiques mais sans réduction des rendements (Rott *et al.*, 1995) est classé comme génotype tolérant (Rott *et al.*, 1994). La ségrégation de sa descendance, évaluée par des tests d'inoculation, fait apparaître des variations continues. De plus, de nombreux QTL de petite taille ont été détectés avec un marquage couvrant environ 50% du génome. Ces résultats suggèrent un déterminisme de la résistance très polygénique dans cette variété et une absence probable de facteur majeur.

Concernant l'héritabilité de la résistance à l'échaudure des feuilles, Egan (1972) fait état d'estimations relativement élevées. Ces résultats sont néanmoins fondés sur l'observation d'un petit nombre de combinaisons génétiques correspondant à du matériel interspécifique datant d'une quarantaine d'années. Plus récemment, une équipe étudiant un dispositif di-allèle avec 6 génotypes a montré des valeurs d'héritabilité de la résistance (au sens strict) modérées ($h^2=0,57$) (Bressiani *et al.*, 2007).

I.4.3 La feuille jaune

a. Impact économique, agent causal

En 2000, la maladie de la feuille jaune était recensée dans 35 régions productrices de canne dans le monde (Lockhart and Cronjé, 2000). Elle provoque chez les cultivars sensibles une réduction des rendements agronomiques et sucriers (Vega *et al.*, 1997 ; Viswanathan, 2002 ; Rassaby *et al.*, 2003 ; Lehrer *et al.*, 2009) associée à une accumulation du saccharose dans les feuilles (Gonçalves *et al.*, 2005 ; Fontaniella *et al.*, 2003).

La maladie de la feuille jaune est causée par le virus SCYLV (Sugarcane Yellow Leaf Virus) appartenant à la famille des *Luteoviridae* (Maia *et al.*, 2000). L'étude de sa séquence a montré que ce virus est issu de recombinaisons entre plusieurs virus d'origines différentes (Moonan *et al.*, 2000 ; Smith *et al.*, 2000). Il a finalement été classé dans le genre *Polerovirus* (D'Arcy et Dommier, 2005). La présence du SCYLV peut être diagnostiquée par immuno-empreintes (Scagliusi and Lockhart, 2000) ou par RT-PCR (Comstock *et al.*, 1998).

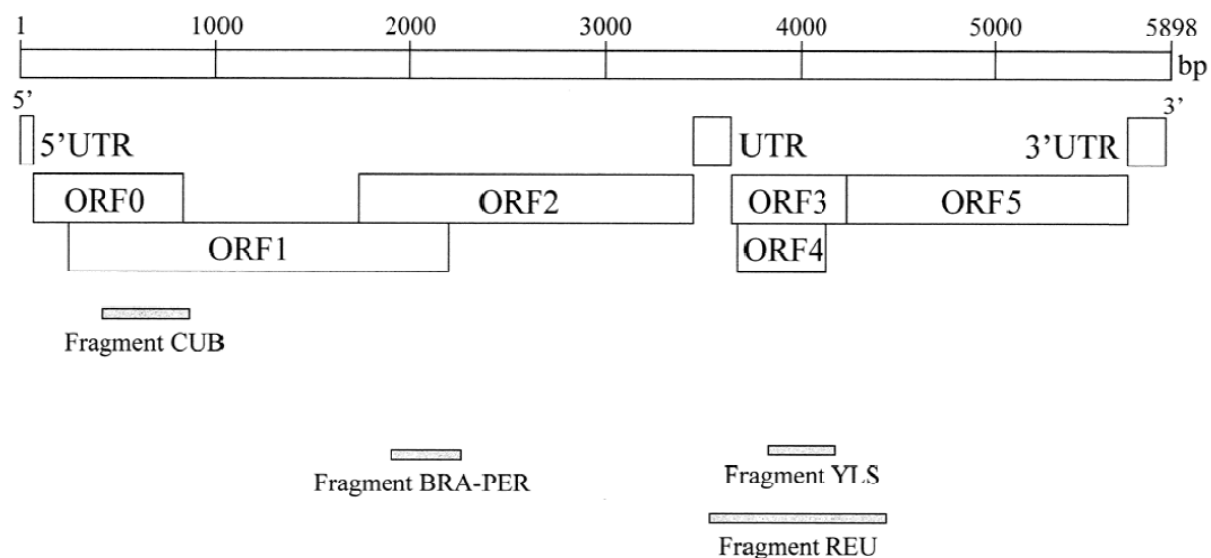


Figure 10 : Organisation du génome du SCYLV et localisation des fragments RT-PCR permettant de différencier les génotypes CUB, BRA-PER et REU. UTR: région transcrite non traduite (UnTranslated Region), ORF : cadre de lecture ouvert (Open Reading Frame) (Abu Ahmad *et al.*, 2006b).



Figure 11 : Symptômes de la maladie de la feuille jaune causée par le SCYLV a) Jaunissement b) Rougissement de la nervure centrale (Photos : J.H. Daugrois, CIRAD).

Le SCYLV est un virus à ARN simple brin composé de 6 ORFs (Open Reading Frame) caractéristiques. La fonction des protéines codées par les 6 ORF a été étudiée chez plusieurs *Luteoviridae*. L'ORF 0 est impliqué dans la suppression des défenses de la plante (inhibition des « Post Transcriptionnal Gene Silencing ») (Pfeffer *et al.*, 2002), les ORF 1 et 2 sont impliquées dans la réplication du virus (Brault *et al.*, 2001), les ORF 3 et 4 codent pour des protéines impliquées dans le déplacement du virus dans la plante, respectivement la protéine de Capside (CP) et la protéine de mouvement (MP) (Brault *et al.*, 2001). Enfin, l'ORF 5 code pour une protéine intervenant dans la traduction de l'ARN, impliquée dans le mouvement du virus dans le vecteur, et indispensable à l'interaction virus-vecteur (Terauchi *et al.*, 2003 ; Brault *et al.*, 2003 et 2005) (**Figure 10**).

b. Expression des symptômes

Le symptôme caractéristique de la maladie est un jaunissement de la nervure centrale sur la face inférieure des feuilles, pouvant évoluer vers un rougissement voire une nécrose se développant à partir de l'extrémité du limbe (Lockhart and Cronje, 2000) (**Figure 11**). Cette symptomatologie n'est toutefois pas suffisante pour diagnostiquer la maladie. Des déficits hydriques peuvent également causer des symptômes semblables en l'absence de virus (Rassaby *et al.*, 2003). Par ailleurs il existe souvent une infection virale sans qu'il y ait de symptômes foliaires (Comstock *et al.*, 1998 ; Izaguirre-Mayoral *et al.*, 2002). En conséquence, seul un diagnostic de laboratoire permet de détecter de façon certaine la présence de la maladie. L'expression des symptômes peut être favorisée par des stress hydriques (excès et déficit) ou bien thermiques (Rott *et al.*, 2008).

En Guadeloupe, le SCYLV est présent sur toute la sole cannière (Edon-Jock *et al.*, 2007). Une parcelle de clones sensibles nouvellement plantée peut être très rapidement infectée si les conditions sont favorables à l'arrivée de populations de pucerons ailés (Daugrois *et al.*, 2011).

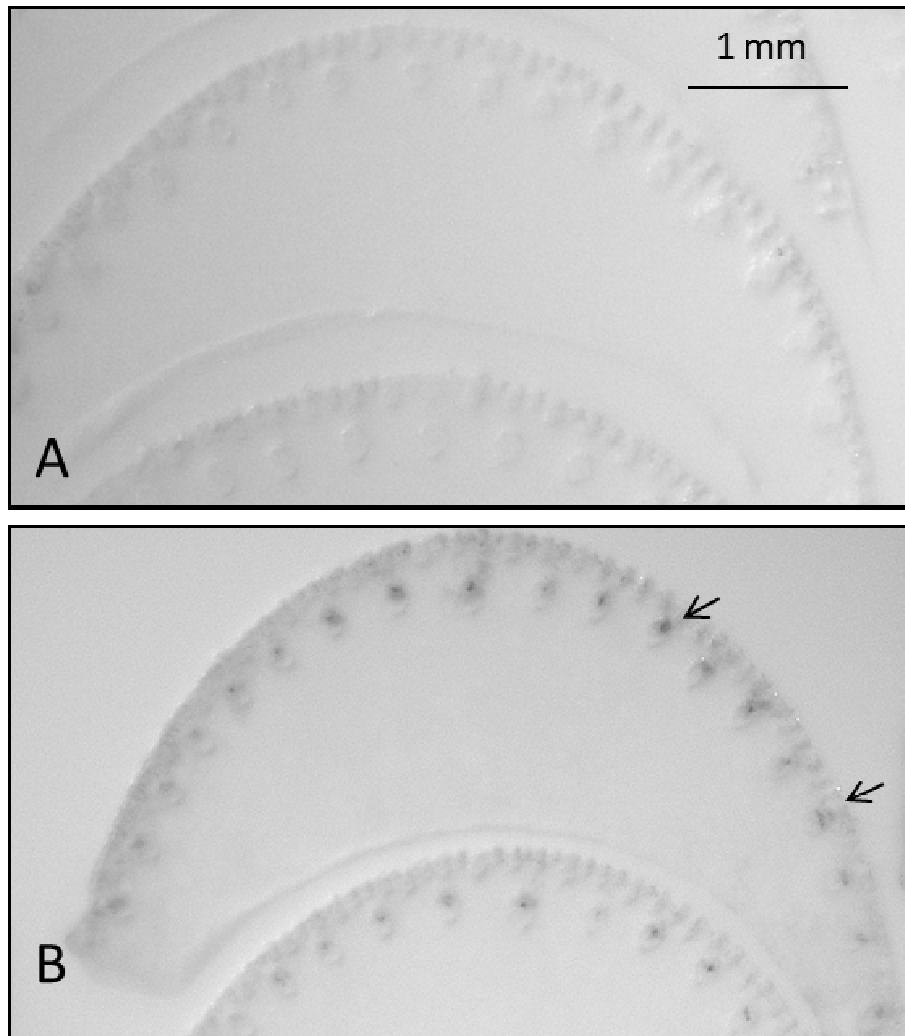


Figure 12 : Empreintes sérologiques (TBIA – Tissue Blot Immuno Assay) de nervure centrale de feuilles de canne révélant la présence du SCYLV dans les cellules compagnes du phloème (cf. protocole annexe 1). A) Feuille saine B) Feuille très contaminée, la coloration bleue de tous les faisceaux vasculaires révèle la présence du SCYLV (Photos C. Edon, CIRAD).

c. Transmission de la maladie

Le virus peut être propagé par la plantation de boutures infectées (Lockhart and Cronjé, 2000). Comme les autres virus de la famille des *Luteoviridae*, le SCYLV est transmis par puceron sur le mode circulant (ingestion, passage de la barrière intestinale, hémocèle, glandes salivaires), non multipliant (absence de réplication des particules virales dans le puceron). Chez les plantes infectées les particules virales visibles dans des empreintes sérologiques sont localisées dans les cellules compagnes du phloème (**Figure 12**) (Vega *et al.*, 1997). Le puceron acquiert le virus en s'alimentant dans un phloème infecté. L'efficacité de l'ingestion du virus par le vecteur dépend notamment du temps passé par le puceron à s'alimenter sur la plante et de la concentration du virus. L'aptitude à l'acquisition du virus et sa circulation à l'intérieur du puceron nécessaires à sa retransmission dépendent d'interactions complexes entre le vecteur et le virus (Gray and Gildow, 2003), dont un des aspects les plus importants réside dans la spécificité de la reconnaissance du virus nécessaire au franchissement de la barrière intestinale et des glandes salivaires (Yang *et al.*, 2008).

Plusieurs espèces de pucerons peuvent transmettre le SCYLV : *Melanaphis sacchari*, *Rhopalosiphum maidis*, *Rhopalosiphum rufiabdominalis* et *Certovacuna lanigera*, (Rott *et al.*, 2008 ; Scagliusi and Lockhart, 2000). Toutefois *M. sacchari* est un vecteur plus efficace que les deux espèces de *Rhopalosiphum* (Schenck and Lehrer, 2000). C'est le principal vecteur présent à la Réunion et en Guadeloupe sur la canne à sucre (Rassaby *et al.*, 2004 ; Daugrois *et al.*, 2011). En région tropicale, comme la quasi-totalité des espèces d'aphides, *M. sacchari* possède un comportement anholocyclique : il se multiplie parthénogénétiquement sur son hôte toute l'année. Lorsqu'il y a surpopulation, dépérissement ou vieillissement de la plante hôte, des femelles ailées apparaissent au sein des colonies, permettant ainsi la dissémination des pucerons vers de nouvelles cannes à sucre (Edon, 2008).

Une récente étude de la diversité génétique dans le groupe des *M. sacchari* menée à partir d'échantillons collectés dans 14 régions de par le monde et basée sur 10 marqueurs microsatellites, a permis de mettre en évidence une structuration mondiale en 4 groupes génétiques majeurs, à l'intérieur desquels il existe une très faible diversité génétique. Cette étude qui confirme l'absence de reproduction sexuée fait apparaître que les populations de pucerons de Guadeloupe et de la Réunion appartiennent à un même et unique groupe. Ce

groupe est le plus répandu à l'échelle mondiale (Amérique du Sud, Caraïbes, Océan Indien) et contient un génotype prépondérant suspecté d'être en expansion (Fartek *et al.*, b, soumis).

Le SCYLV est un virus spécifique de la canne à sucre. Aucune graminée adventice ne l'héberge (Schenck and Lehrer, 2000 ; Komor, 2011). Il semble que le virus soit transmissible au blé et à l'avoine (Schenck and Lehrer, 2000).

d. Diversité génétique du pathogène

Plusieurs analyses de diversité génétique du SCYLV ont permis de différencier 6 génotypes viraux qui ont été dénommés par le lieu où ils ont été détectés pour la première fois : REU (Réunion), BRA (Brésil), PER (Pérou), CUB (Cuba), IND (Inde) et CHIN1 (Chine) (Abu Ahmad *et al.*, 2006a et 2007 ; Viswanathan *et al.*, 2008 ; Wang and Zhou, 2010). Des amorces spécifiques ont été définies par Abu ahmad *et al.* (2006b) pour différencier les 4 premiers génotypes identifiés par RT-PCR. A l'exception des génotypes BRA et PER qui ne sont pas différenciables, des amorces spécifiques amplifiant des zones d'ORF distinctes ont été mises au point pour la détection de chaque génotype viral (BRA-PER, CUB et REU) (**Figure 10**). Le génotype BRA-PER est très répandu dans le monde à la différence des génotypes REU et CUB (Abu Ahmad *et al.*, 2006b).

Ces trois génotypes existent en Guadeloupe, alors que seuls les deux premiers sont présents à la Réunion. Une étude comparative, portant sur un même petit panel de 8 variétés respectivement inoculées dans chacune des deux îles, à partir d'isolats correspondant à chacun des deux même génotypes viraux (BRA-PER et REU), a montré des incidences du virus différentes entre les deux îles (Abu Ahmad *et al.*, 2007). Ce résultat associé à d'autres considérations illustre l'existence de contextes épidémiologiques nécessairement spécifiques à chaque lieu. Un génotype viral donné affectant une même variété ne montrera pas nécessairement la même capacité d'agressivité (aptitude à l'infection de la plante) ni la même virulence (intensité de la maladie) d'un lieu à l'autre.

e. Sources de résistance chez l'hôte et nature du déterminisme

Actuellement, il n'existe pas de traitement satisfaisant contre la maladie de la feuille jaune. L'assainissement des boutures par culture *in vitro* de méristèmes est efficace, mais longue et coûteuse, et les variétés sensibles sont rapidement réinfectées au champ (Rassaby *et al.*, 2004). Les traitements chimiques contre le puceron vecteur sont techniquement difficiles à mettre en œuvre, coûteux sur les plans économique et environnemental, et propices au développement de résistances aux insecticides chez le puceron. L'amélioration génétique de la résistance de la canne à sucre, à la fois contre *M. sacchari* et contre le SCYLV, constitue la principale stratégie de lutte méritant d'être privilégiée.

- **Résistance à *M. sacchari*.**

Les pucerons sont des ravageurs de nombreuses cultures, notamment à cause de leur cycle de vie très court et de leur taux de reproduction élevé, et parce qu'ils sont vecteurs de nombreuses maladies. Des sources de résistance aux aphides ont été identifiées chez de nombreuses espèces cultivées. Ces sources de résistance sont souvent spécifiques d'une espèce de puceron, et reposent le plus souvent sur un nombre de gènes et d'allèles limité (Dogimont *et al.*, 2010).

Deux mécanismes distincts interviennent dans la résistance aux pucerons : l'antixénose (ou non-préférence), qui repose sur la tendance de l'insecte à contaminer une autre plante, et l'antibiose, qui repose sur la moindre capacité de l'insecte à vivre et se multiplier sur la plante (Dogimont *et al.*, 2010).

Akbar *et al.* (2010, 2011) ont identifié des sources de résistance partielle à *M. sacchari* par antibiose chez 3 cultivars de canne à sucre. Les auteurs n'ont pas mis en évidence de mécanismes d'antixénose. L'étude de Fartek *et al.* (a, soumis) a montré que des mécanismes de résistance partielle à un génotype de *M. sacchari*, par antibiose et antixénose existent chez le cultivar étudié, et sont associés à une modification des traits de vie du puceron (taux de survie et de reproduction du puceron réduit), ainsi que de son comportement alimentaire (augmentation du délai de salivation et absence d'ingestion de sève pendant les premières heures suivant l'arrivée des pucerons).

- Résistance au SCYLV

Parmi les hybrides contemporains, il existe de nombreuses variétés relativement résistantes au SCYLV qui montrent un niveau d'incidence de la maladie faible voir nul (Victoria *et al.*, 2005 ; Komor, 2011). Par ailleurs, parmi les ressources génétiques ancestrales, une grande majorité d'accessions *S. spontaneum* s'avère résistante au SCYLV, alors que la majorité des clones *S. officinarum* est sensible (Comstock *et al.*, 2001b ; Komor, 2011). Il existe donc un potentiel de résistance intéressant qui peut être exploité en sélection.

Jusqu'à présent peu de descendance impliquant un géniteur résistant ont été testées pour leur ségrégation au SCYLV. L'étude de descendance biparentale permettrait pourtant d'identifier d'éventuels gènes de résistance majeurs susceptibles d'être valorisés dans les programmes d'amélioration.

Komor (2011), en étudiant le croisement d'un clone *S. robustum* résistant par un clone *S. officinarum* sensible, a obtenu une ségrégation agronomique très fortement déséquilibrée vers les résistants (85 % résistants : 15 % sensibles) qu'il attribue à un caractère de « dominance ».

Victoria *et al.*, 2005 ont étudié un croisement *résistant x sensible* entre deux variétés locales. Ces auteurs observent une proportion (stable sur trois répétitions) d'environ 71 % de descendants résistants et de 29 % de descendants sensibles. Ces proportions obtenues sur une population totale de 148 individus s'avèrent parfaitement compatibles (test du chi-2 très significatif) avec l'hypothèse de l'existence de deux gènes majeurs présents chez le parent résistant.

Enfin récemment, l'analyse du croisement 'MQ76-53 x R570' a permis de mettre en évidence la présence d'un gène majeure de résistance chez MQ76-53 (Costet *et al.*, soumis).

I.5 Génome de la canne

I.5.1 Génome des variétés modernes

Les cultivars modernes sont des clones polyploïdes d'origine bispécifique (*S. officinarum* x *S. spontaneum*) possédant un génome particulièrement complexe, aneuploïde,

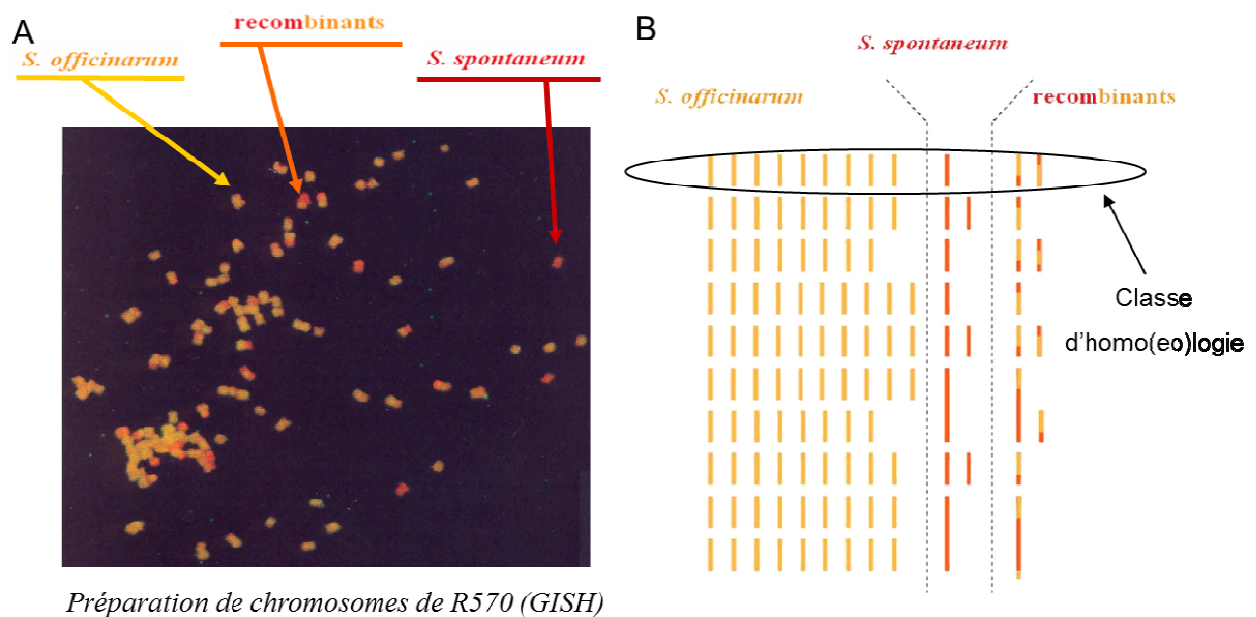


Figure 13 : Structure du génome du cultivar R570 ($2n = \text{ca } 115$).

A) Hybridation génomique *in situ* (GISH) de chromosomes du cultivar R570 en utilisant l'ADN génomique de *S. officinarum* (marqué en jaune) et l'ADN génomique de *S. spontaneum* (marqué en rouge) (D'Hont *et al.*, 1996).

B) Représentation schématique du génome de R570. Chaque trait vertical représente un chromosome. Les chromosomes issus du parent *S. officinarum* sont représentés en jaune et ceux du parent *S. spontaneum* en rouge. Les chromosomes présents sur une même ligne appartiennent au même groupe d'hom(é)ologie. Les chromosomes présents sur une même colonne appartiennent au même génome haploïde (D'Hont, 2005).

constitué de $2n = 100$ à 130 chromosomes répartis en 10 classes d'homologie. Chaque chromosome de base est présent en 10 à 14 copies hom(é)ologues en moyenne (D'Hont *et al.*, 1996 ; Grivet and Arruda, 2001). La double origine spécifique du génome des cultivars modernes a été étudiée par D'Hont *et al.* (1996) par hybridation génomique *in situ* (GISH) (**Figure 13**). La composition génétique des cultivars modernes est variable d'un clone à l'autre, notamment selon les schémas d'introgessions pratiqués (*S. spontaneum* dans *S. officinarum*) et le cytotype du clone *S. spontaneum* impliqué dans le croisement interspécifique de départ. La technique de GISH, qui permet de révéler avec différents fluorochromes les fractions du génome issues des deux espèces ancestrales, a montré qu'environ 73 à 85 % des chromosomes des cultivars modernes sont entièrement d'origine *S. officinarum*, que 10 à 23 % d'entre eux sont uniquement d'origine *S. spontaneum* et que 8 à 13 % des chromosomes sont des recombinants entre les deux espèces ancestrales (D'Hont *et al.*, 1996 ; Piperidis *et al.*, 2010). Cette technique de cytogénétique qui permet de mesurer le poids relatif des sous-génomes ancestraux dans les clones issus de « nobilisation » a aussi permis de lever des erreurs de classifications d'individus entre le groupe interspécifique et l'espèce noble (*S. officinarum*) (Piperidis *et al.*, 2010).

La transmission de $2n$ chromosomes d'origine *S. officinarum* pour n chromosomes d'origine *S. spontaneum* lors du premier croisement interspécifique et du premier backcross a contribué à une diminution rapide du poids relatif du sous-génome *S. spontaneum* dans les génotypes hybrides (Bremer, 1961 ; Bhat and Gill, 1985 ; Piperidis *et al.*, 2010). Toutefois en dépit du poids relativement modeste de l'espèce sauvage dans le génome des hybrides, une partie importante de la diversité génétique de ces derniers est attribuable à l'espèce sauvage en raison du polymorphisme sensiblement supérieur que cette espèce renferme comparativement à l'espèce noble (Lu *et al.*, 1994).

Chez les hybrides la méiose n'est pas complètement régulière. Même si 90 à 95 % des appariements se font sous forme de bivalents, on dénombre toujours quelques rares univalents ou multivalents (Burner and Legendre, 1993). Ces petites irrégularités ont sans doute trait aux différences de structure qui distinguent les deux sous-génomes composant les hybrides ($x=8$ pour *S. officinarum*, $x=10$ pour *S. spontaneum*). Contrairement à ce que la formation prépondérante de bivalents pourrait suggérer, le mode d'appariement des chromosomes à la méiose n'est pas complètement de type disomique chez les hybrides. Il montre une situation complexe combinant appariements aléatoires entre certains

Tableau 3 : Cartes génétiques de la canne à sucre.

Espèce	Clone / cultivar	Nombre de marqueurs	Type de marqueurs ⁽¹⁾	GC/GH ⁽²⁾	Taille (cM)	Référence
<i>S. spontaneum</i>	SES 208	208	AP-PCR	42/-	1500	Aljanabi <i>et al.</i> , 1993
		216	AP-PCR	44/-	-	Da Silva <i>et al.</i> , 1993
		527	AP-PCR, RFLP	64/8	-	Da Silva <i>et al.</i> ,1995
<i>S. spontaneum</i>	IND81-146	248	RFLP	69/10	2063	Ming <i>et al.</i> , 1998
	PIN84-1	182	RFLP	69/7	1303	
<i>S. officinarum</i>	Green German	270	RFLP	72/10	2034	
	Muntok Java	206	RFLP	72/9	1443	
<i>S. officinarum</i>	LA Purple	345	AP-PCR, RFLP, AFLP	74/-	-	Guimaraes <i>et al.</i> , 1999
<i>S. robustum</i>	Mol5829	301	AP-PCR, RFLP, AFLP	65/-	-	
<i>S. officinarum</i> x <i>S. spontaneum</i>	Carte consensus	982	RFLP	289/13	-	Ming <i>et al.</i> , 2002a
<i>S. officinarum</i> x <i>S. spontaneum</i>	La Striped x SES 147B	650	AFLP, SRAP, TRAP	49/-	1491	Alwala <i>et al.</i> , 2008
<i>S. spp</i>	SP701006	94	RFLP	25/8	-	D'Hont <i>et al.</i> , 1994
	R570	408	RFLP	96/10	-	Grivet <i>et al.</i> , 1996
		608	RFLP, AFLP, SSR	162/8	-	
		939	AFLP	120/10	5849	Hoarau <i>et al.</i> , 2001
		1087	RFLP, AFLP, SSR	128/7	-	Rossi <i>et al.</i> , 2003
		1123	RFLP, AFLP, SSR	128/7	7675	Raboin <i>et al.</i> , 2006
	Q165	1074	AFLP, SSR, RAF	136/8	9058	Aitken <i>et al.</i> , 2005
	SP80-180 x SP80-4966	1118	RFLP, AFLP, SSR	131/12	2602,4	Garcia <i>et al.</i> , 2005
	Q117 x MQ77-340	400	AFLP, SSR	101/-	3582	Reffay <i>et al.</i> , 2005
	MQ 76-53	536	RFLP, AFLP, SSR	105/7	4329	Raboin <i>et al.</i> , 2006
	M 134/75	559	AFLP, SSR	95/-	6200	Aljanabi <i>et al.</i> , 2007
	SP80-180 x SP80-4966	2303	RFLP, AFLP, SSR, EST-RFLP, EST-SSR	192/14	6261,1	Oliveira <i>et al.</i> , 2007
	LCP 85-384	1111	AFLP, SSR, TRAP	108/9	5617	Andru <i>et al.</i> , 2011

(1) AP-PCR: Arbitrarily Primed PCR RAF: randomly amplified fingerprints (2)GC: groupe de coségrégation GH: groupe d'homologie

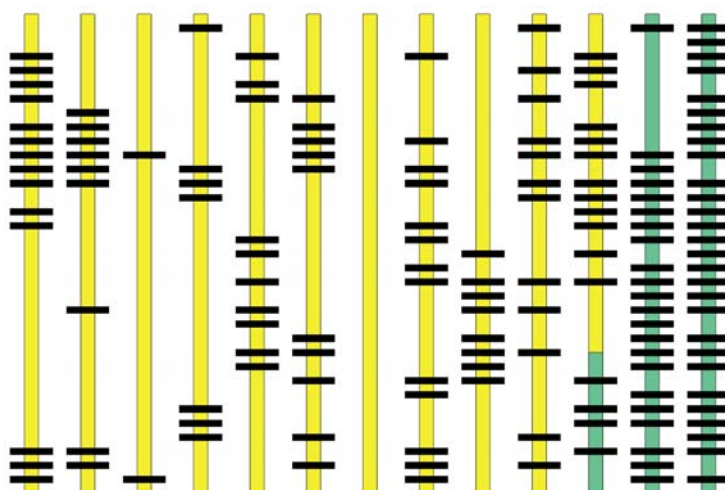


Figure 14 : Représentation schématique d'une classe d'homéologie caractéristique d'un cultivar de canne à sucre. Les chromosomes et régions chromosomiques issus de *S. officinarum* sont représentés en jaune, ceux issus de *S. spontaneum* sont représentés en vert. Les traits noirs horizontaux représentent les mutations spécifiques (portées par un seul chromosome), accessibles au marquage par des marqueurs simplex. (D'après Grivet and Arruda, 2001).

chromosomes et appariements préférentiels plus ou moins forts entre d'autres chromosomes (Grivet *et al.*, 1996 ; Hoarau *et al.*, 2001 ; Jannoo *et al.*, 2004).

Une analyse de diversité réalisée par Jannoo *et al.* (1999b), au moyen de 12 sondes RFLP et portant sur 53 clones représentatifs de *S. officinarum* et sur 109 cultivars modernes, a montré que ces derniers renferment plus de 80% de la diversité allélique existant chez *S. officinarum*. Compte tenu du faible nombre de fondateurs *S. officinarum* utilisés en croisements (Arceneaux, 1965), cette relative bonne représentation de la diversité de l'espèce domestiquée dans les cultivars modernes s'explique par la modeste diversité génétique observée dans les ressources *S. officinarum* jusqu'à présent étudiées.

I.5.2 Cartographie génétique

a. Cartes génétiques

La canne, particulièrement les cultivars modernes à cause de leur origine hybride entre deux espèces polyploïdes, est la plante qui possède le génome le plus complexe qui ait été étudié à ce jour. Compte tenu de la ploïdie élevée, les marqueurs sont toujours interprétés en codage dominant (y compris les marqueurs RFLP et SSR locus-spécifiques). La cartographie est uniquement basée sur l'utilisation de marqueurs simplex¹ (Wu *et al.*, 1992). Elle consiste à définir des groupes de coségrégation (GC) de marqueurs simplex. Les marqueurs duplex ne peuvent pas être utilisés pour la cartographie, mais peuvent servir à identifier des GCs appartenant au même groupe d'homologie (Grivet, 1995).

Plusieurs cartes génétiques de divers clones ancestraux (*S. spontaneum*, *S. officinarum* et *S. robustum*) et de cinq cultivars modernes ont été produites à l'aide de marqueurs RAPD, RFLP, AFLP, SSR (**Tableau 3**). En raison du nombre limité de marqueurs simplex dans certaines zones du génome, aucune de ces cartes ne couvre l'intégralité du génome. Comme l'illustre la **Figure 14**, les chromosomes hybrides hérités de l'espèce

¹ Marqueur simplex (ou marqueur simple dose) : présent en une seule copie chez un des parents et absent chez l'autre, et qui ségrégera donc en ratio 1 présent:1 absent dans les gamètes.

Tableau 4 : Principales études QTL réalisées chez la canne à sucre.

Taille de population	Croisement ⁽¹⁾	Nombre de marqueurs	Type de marqueurs ⁽²⁾	Caractère	NQTA (Nm _q) ⁽³⁾	r ² individuel (%) ⁽⁴⁾	r ² global (%) ⁽⁴⁾	r ² moyen (%) ⁽⁴⁾	Seuil	Référence
100	SP80-180 (Cv) x SP80-4966 (Cv)	47 + 221	EST + RFLP	Taux de fibres	- (15)	3,88-10,9	28,14-29,73			
				Rendement agronomique	- (22)	3,87-18,12	18,96-22,96			
				Pol	- (27)	3,98-12,6	18,62-19,76		p≤0,05	Pinto <i>et al.</i> , 2010
				TSH (Tonnes de sucre par Ha)	- (25)	4,03-20,1	25,01-33,34			
264	Green german (<i>Off</i>) x IND81-146 (<i>Spont</i>)	243+232	RFLP	Rendement en sucre	14 (-)	3,6-20,6	65,5	4,7		
239	Muntok java (<i>Off</i>) x PIN84-1 (<i>Spont</i>)	122+138	RFLP	Rendement en sucre	22 (-)	3,6-20,6	68,3	3,1	p≤0,003	Ming <i>et al.</i> , 2001
264	Green german (<i>Off</i>) x IND81-146 (<i>Spont</i>)	735	RFLP	Rendement en sucre	3 (-)	5-10,7	18,4	6,1		
				Pol	2 (-)	5,1-8,8	18,5	9,3		
				Poids des tiges	10 (-)	5,9-13,3	62,7	6,3		
				Nombre de tiges	2 (-)	5,3-15	13,9	7		
				Teneur en fibres	19 (-)	5,6-13,2	60,6	3,2		
				Cendres	7 (-)	5,2-11,2	39,1	5,6		
239	Muntok java (<i>Off</i>) x PIN84-1 (<i>Spont</i>)	735	RFLP	Rendement en sucre	7 (-)	4,6-15,4	30,2	4,3	p≤0,003	Ming <i>et al.</i> , 2002b
				Pol	12 (-)	4,1-15,4	39,9	3,3		
				Poids des tiges	25 (-)	4,7-16,2	71,6	2,9		
				Nombre de tiges	1 (-)	6,1	6,5	6,5		
				Teneur en fibres	1 (-)	7	7,0	7		
				Cendres	13 (-)	3,8-12,4	41,4	3,2		
264	Green german (<i>Off</i>) x IND81-146 (<i>Spont</i>)	735	RFLP	Taille des tiges	- (3)	4,6-22,6	16,3			
				Floraïson	- (9)	5-8,4			p≤0,003	Ming <i>et al.</i> , 2002c
239	Muntok java (<i>Off</i>) x PIN84-1 (<i>Spont</i>)	735	RFLP	Taille des tiges	- (53)	4,6-22,6	-			
295	R570 (Cv) x R570 (Cv)	1000	AFLP	Taille des tiges	12 (27)	2-6	30-35	2,7		
				Nombre de tiges	8 (31)	3-7	36-40	4,8		
				Diamètre des tiges	9 (38)	3-7	44-51	5,3	p≤0,005	Hoarau <i>et al.</i> , 2002
				Brix	9 (12)	3-5	46-55	5,6		
108	Q117 (Cv) x 74C42 (Cv)	258	RFLP, RAF	Nombre de tiges	(16)	-	-			
				"Suckering"	(14)	-	-		p≤0,01	Jordan <i>et al.</i> , 2004
232	Q117 (Cv) x MQ77-340 (Cv)	400	AFLP, SSR	Pol	8(11)	3-7	-			
				Brix	8(13)	3-6	-			
				CCS	7(9)	3-7	-			
				Fibre	1(1)	2-3	-		p≤0,001	Reflay <i>et al.</i> , 2005
				Poids des tiges	6(7)	2-7	-			
				Rendement agronomique	1(1)	3	-			
				Rendement en sucre	3(3)	3	-			
230	Q165 (Cv) x IJ76-514 (<i>Off</i>)	1000	AFLP, SSR	Brix	13 (16)	3-9	12-29	1,6	p≤0,001	Aitken <i>et al.</i> , 2006
				Pol	11(17)	4-8	13-31	2		
227	M 134/75 (Cv) x R570 (Cv)	666	AFLP, SSR	Résistance au 'Yellow spot'	2 (5)	7,7-17,6			LOD >3	Al Janabi <i>et al.</i> , 2007
227	Q165 (Cv) x IJ76-514 (<i>Off</i>)	1000	AFLP, SSR	Poids des tiges	10 (13)	4-10	43-57	5		
				Diamètre des tiges	11 (16)	4-7	43-60	4,7		
				Nombre de tiges	8 (10)	4-9	43-54	6,1	p≤0,001	Aitken <i>et al.</i> , 2008
				Taille des tiges	8 (11)	4-7	27-29	3,5		
				Biomasse totale	13 (15)	4-7	29-59	3,4		
147	R570 (Cv) x R570 (Cv)	1405	AFLP, SSR, RFLF	Résistance au borer <i>Chilo sacchariphagus</i>	7 (21)	7-12	42 (6 QTA)	7	p≤0,001	Nibouche <i>et al.</i> , 2011

(1) *Off*: *S. officinarum* ; *Spont* : *S. spontaneum* ; Cv: Cultivar (2) RAF: Radiolabeled amplified fragment (3) NQTA(Nm_q) : Nombre de QTA associés au caractère (Nombre de marqueurs détectés) (4) r²: part de variation expliquée par le marquage

sauvage ont tendance à être mieux couverts que les chromosomes originaires de *S. officinarum* en raison d'un nombre plus abondant de marqueurs simplex.

La carte de référence du CIRAD développée depuis plusieurs années a été réalisée à l'aide d'une population issue de l'autofécondation de R570 (Grivet *et al.*, 1996 ; Hoarau *et al.*, 2001 ; Rossi *et al.*, 2003 ; Raboin *et al.*, 2006). Elle contient actuellement près de 2000 marqueurs de divers types (SSR, RFLP, AFLP et DArT) qui forment 151 GCs correspondant à 7 groupes d'hom(e)ologies (HG) (Raboin *et al.*, 2006 et données non-publiées). Cette carte couvre environ 9 000 cM, soit un peu plus de 50 % du génome, dont la taille a été estimée à 17 000 cM (Hoarau *et al.*, 2001), pour un génome de 10 000 Mpb (D'Hont, 2005).

b. Etudes QTL

Seuls trois gènes majeurs ont été identifiés chez la canne : deux gènes de résistance à la rouille et un gène contrôlant la couleur des tiges (Daugrois *et al.*, 1996 ; Asnaghi *et al.*, 2004 ; Raboin *et al.*, 2006).

Quelques études de détection d'allèles à effet quantitatif (QTA), concernant la résistance à quelques maladies et les composantes du rendement agronomique ou sucrier, ont été réalisées sur plusieurs accessions de canne correspondant à des génotypes élités ou à des clones ancestraux. Ces études, recensées dans le **Tableau 4**, ne sont pas toutes comparables entre elles car elles varient en termes de taille de population, de densité de marquage, de seuils statistiques de détection des QTA. Ces études ont permis de révéler que la majeure partie des caractères d'intérêt chez la canne à sucre sont sous le contrôle de nombreux QTA à effet individuel faible. Hormis quelques exceptions, la part de variation expliquée par des marqueurs (R^2) dépasse rarement 10 à 12 %. Ces R^2 relativement modestes s'expliquent par l'existence de nombreux allèles à chaque locus, conséquence des niveaux de ploïdie et d'hétérozygotie élevés (Hoarau *et al.*, 2002). L'effet cumulé des QTA (R^2 global) varie largement selon l'effort de génotypage, le matériel végétal étudié et les caractères considérés.

Concernant les caractères liés aux rendements (agronomique et sucrier), les R^2 individuels des QTA sont globalement plus faibles dans les croisements entre fonds cultivés (cultivars interspécifiques ou clones *S. officinarum*) (2 à 10 %), que dans les croisements

Tableau 5 : Quelques études d'associations réalisées chez plusieurs espèces végétales.

Espèce	Taille du génome haploïde (Mpb)	Ploidie	Taille de population (type) ⁽¹⁾	Marquage	Etendue du DL (cM)	Caractère	Structure ⁽²⁾	Nombre d'associations	R ² individuel (%)	R ² global (nombre de marqueurs) (%)	Référence
<i>Arabidopsis thaliana</i>	115	2n=10	95 (E)	876 Séquences		Date de floraison	-	55			Zhao <i>et al.</i> , 2007
							Q ^(SA) +K	4			
							K*	6			
							Q ^(SA) +K*	4			
							Q*+K	5			
							Q*+K*	5			
			275 (E)	319 SNP (51 gènes candidats)		Date de floraison (Jours courts)	Q ^(SA) +K*	0			Ehrenreich <i>et al.</i> , 2009
						Nombre de feuilles (Jours courts)		0			
						Date de floraison (Jours longs)		10	3-12		
						Nombre de feuilles (Jours longs)		1	2		
<i>Avena sativa</i> L.	11300	2n=6X=42	114 (Cv)	77 AFLP		Rendement agronomique et qualité du grain (9 caractères)	-	55	6-34		Achleitner <i>et al.</i> , 2008
							Q*(axe 1-4)	27	5-25		
							Q ^(SA) +K	40	3-23		
<i>Brassica napus</i> L.	560	2n=4X=38	128 (Cv)	227 SSR, SCAR, SSR _{spé}		Résistance à <i>Lepstosphaeria maculans</i>	Q ^(SA)	49			Jestin <i>et al.</i> , 2011
							K	49	3-12		
							Q ^(SA) +K	23			
<i>Gossypium</i>	2200-3000	2n=4X=52	260 (P)	202 SSR		Qualité de la fibre (6 caractères)	-	59			Zeng <i>et al.</i> , 2009
			334 (Cv)	202 SSR	0-25	Qualité de la fibre (6 caractères)	Q ^(SA) +K	39			
<i>Hordeum vulgare</i>	4800	2n=2X, 4X, 6X=14, 28, 42	220 (Col)	3 gènes candidats		Date de floraison	Q ^(SA) +K	24	9,3-32,4	51(2)	Stracke <i>et al.</i> , 2009
							Q*+K		3,4-22,4	48,8(2)	
			131 (Col)	83 SNP + 12 gènes candidats	1-5,5	Resistance au BYMV	Q ^(SA)	36	-	6	Stracke <i>et al.</i> , 2007
			103 (Cv)	123 SSR	0-25	rendement et kernel	Q ^(SA) +K	25	10,2-32,6	51	Agrama <i>et al.</i> , 2007
<i>Oryza sativa</i>	450	2n=2X=24	116 (RIL-HD)	118 (PCR-RGA, PCR-RFLP, gènes candidats)		Résistance à <i>Xanthmonas Oryzae</i>	-	6	8,4-14,3		Ramalingam <i>et al.</i> , 2003
						Résistance au Brown Plant Hopper	-	9	11,3-20,6		

(1) AP: Population d'association Cv: cultivar Col: Collection E: écotypes HD: haploïdes doublés LE: lignées élite NAP: 'Nested' association population (population d'association avec admixture/sturcturation) P: Polycross (population issue de nombreuses générations de croisements interspécifiques suivies de nombreuses générations d'autofécondation) RIL: lignées recombinantes

(2) Q(SA): Structure d'après Pritchard *et al.* (2000a) Q*: structure d'après méthode ACP (Price *et al.* , 2006; Patterson *et al.* , 2006) Qp: structure d'après données de pedigree K: Kinship d'après Yu *et al.* (2006) K*: Kinship d'après Zhao *et al.* , 2007

espèce cultivée x espèce sauvage (3,8 à 22, 6 %) (**Tableau 4**). Cette différence s'explique par l'utilisation de parents génétiquement moins distants, donnant des ségrégations moins larges des caractères fortement sélectionnés. Les allèles favorables sont mieux « fixés » dans les fonds génétiques cultivés et les allèles à effet fortement négatif ont souvent été éliminés dès les premières étapes de nobilisation ou de sélection (Hoarau *et al.*, 2002 ; Aitken *et al.*, 2006).

Plusieurs études (Hoarau *et al.*, 2002 ; Reffay *et al.*, 2005 ; Aitken *et al.*, 2006 et 2008 ; Pinto *et al.*, 2010) montrent souvent une instabilité relative des détections de QTA entre différents environnements (temporel ou spatial) à des seuils de détection stringents. Toutefois, la répétabilité des détections s'améliore lorsqu'on augmente le risque de première espèce (erreur de type I). Ces résultats s'expliquent par l'existence fréquente d'interactions génotype x environnement et illustrent la sensibilité de la détection de QTA à un effet de seuil statistique dans le contexte de la canne, où le déterminisme de nombreux caractères est de nature probablement d'autant plus quantitative que le niveau de ploïdie est très élevé.

Quelques marqueurs sont associés à plusieurs caractères, et reflètent l'existence de corrélations entre ces derniers. Cela peut être dû à des QTA possédant des effets pléiotropiques ou à l'existence de régions riches en gènes, regroupant plusieurs QTA correspondant à des caractères distincts. L'observation d'interactions entre QTA, par ailleurs quelquefois répétables entre cycles de culture, illustre l'importance potentielle des phénomènes d'épistasie sous-jacents à la variation phénotypique des caractères complexes identifiés (Ming *et al.*, 2002b et c ; Hoarau *et al.*, 2002 ; Aitken *et al.*, 2008 ; Pinto *et al.*, 2010).

Tableau 5 (suite) : Quelques études d'associations réalisées chez plusieurs espèces végétales.

Espèce	Taille du génome haploïde (Mpb)	Ploïdie	Taille de population (type) ⁽¹⁾	Marquage	Etendue du DL (cM)	Caractère	Structure ⁽²⁾	Nombre d'associations	R ² individuel (%)	R ² global (nombre de marqueurs) (%)	Référence
<i>Pennisetium glaucum</i>	400	2n=2X=14	90 (RIL)	306 SSR + AFLP		Taille de la plante	Q ^(SA) +K	1	6,7		Saidou <i>et al.</i> , 2009
						Stade de floraison		8	7,3-9,8		
						Diamètre des tiges		11	3,2-5,6		
						Taille de la hampe florale		16	3,9-7,5		
						Diamètre de la hampe florale		10	3,8-6,6		
<i>Pinus taeda</i> L.	21700	2n=24	498 (AP)	3938 SNP		Résistance à <i>Fusarium circinatum</i>	Q ^(SA)	10	4,7-7,2		Quesada <i>et al.</i> , 2010
<i>Solanum tuberosum</i> L.	850	2n=4X=48	221 (Col)	250 AFLP	0-3	Caractères agro-morpho et qualité du tubercule	-	4-17			D'Hoop <i>et al.</i> , 2008
							K	2-11			
<i>Sorghum</i>	820	2n=20	125 (Cv)	369 SSR + SNP		Taille	Q ^(SA) +K	5		55	Murray <i>et al.</i> , 2009
							Q*+K	5		55	
							Q ^(SA) +K	1		49	
						Brix	Q*+K	1		39	
<i>Triticum aestivum</i>	16000	2n=6X=42	170 (LE)	1644 DArT		Sensibilité des tiges à la rouille brune	Q ^(SA) +K	61	-		Crossa <i>et al.</i> , 2007
						Sensibilité des feuilles à la rouille brune		90	2-12		
						Sensibilité à la rouille jaune		145	1-11		
						Sensibilité au Mildiou		49	3-10		
						Rendement en grains		297	1-10		
<i>Zea maize</i>	2500	2n=20	5000 (NAP)	1106 SNP		Résistance à la rouille	Q ^P	754	1-6	77(29)	Poland <i>et al.</i> , 2011
			102 (RIL)	6 gènes candidats (ae1 ; bt2 ; sh1 ; sh2 ; sugary1 ; waxy1)		Composition du Kernel	Q ^(SA)	3			Wilson <i>et al.</i> , 2004
						Gélatinisation de l'amidon (hiver)		2			
						Gélatinisation de l'amidon (été)		2			
						Teneur en Amylose		3			

(1) AP: Population d'association Cv: cultivar Col: Collection E: écotypes HD: haploïdes doublés LE: lignées élite NAP: 'Nested' association population (population d'association avec admixture/sturcturation) P: Polycross (population issue de nombreuses générations de croisements interspécifiques suivies de nombreuses générations d'autofécondation) RIL: lignées recombinantes

(2) Q(SA): Structure d'après Pritchard *et al.* (2000a) Q*: structure d'après méthode ACP (Price *et al.* , 2006; Patterson *et al.* , 2006) Qp: structure d'après données de pedigree K: Kinship d'après Yu *et al.* (2006) K*: Kinship d'après Zhao *et al.* , 2007

I.6 Génétique d'association

La génétique d'association, basée sur l'**analyse du déséquilibre de liaison (DL)**², vise à identifier les relations entre le polymorphisme génétique et les variations phénotypiques observées dans des populations en ségrégation. Initialement développée en génétique humaine, la génétique d'association est utilisée depuis peu en génétique végétale et animale. Elle consiste en des études d'associations visant à disséquer le déterminisme génétique de caractères utiles qui peuvent faciliter l'identification, dans des ressources génétiques, d'individus susceptibles de porter des allèles d'intérêt. Ces études d'association peuvent être réalisées sur des populations de taille importante, et composées d'individus divers, éventuellement génétiquement liés pour tout ou partie entre eux à des degrés variés.

La littérature récente montre de nombreuses études portant sur un nombre croissant d'espèces végétales. Le **Tableau 5** recense quelques exemples d'études d'associations récentes. On peut y distinguer i) des études effectuées avec un nombre limité de marqueurs visant à rechercher des associations entre différentes formes alléliques de gènes candidats et le phénotype (Ehrenreich *et al.*, 2009 ; Stracke *et al.*, 2007 et 2009; Ramalingam *et al.*, 2003 ; Wilson *et al.*, 2004) et ii) des études effectuées avec de nombreux marqueurs (AFLP, DArT, SNP) balayant l'ensemble du génome, concernant des caractères complexes, qui ont pour but la recherche de marqueurs en déséquilibre de liaisons avec des allèles agronomiques d'intérêt, dispersés à de nombreux loci dans le génome (Zhao *et al.*, 2007 ; Achleitner *et al.*, 2008 ; Zeng *et al.*, 2009 ; Abdurakhmonov *et al.*, 2009 ; Stracke *et al.*, 2007 ; Agrama *et al.*, 2007 ; Ramalingam *et al.*, 2003 ; Saidou *et al.*, 2009 ; Quesada *et al.*, 2010 ; Crossa *et al.*, 2007 ; Jestin *et al.*, 2011 ; Poland *et al.*, 2011). On parle dans ce dernier cas d'approches « genome wide » ou « genome scan ». Compte tenu de l'avènement de marquage haut-débit chez de nombreuses plantes, les études d'associations sur l'ensemble du génome devraient connaître un essor croissant dans les années qui viennent.

² Déséquilibre de liaison, ou déséquilibre d'association gamétique : association non aléatoire d'allèles à différents loci

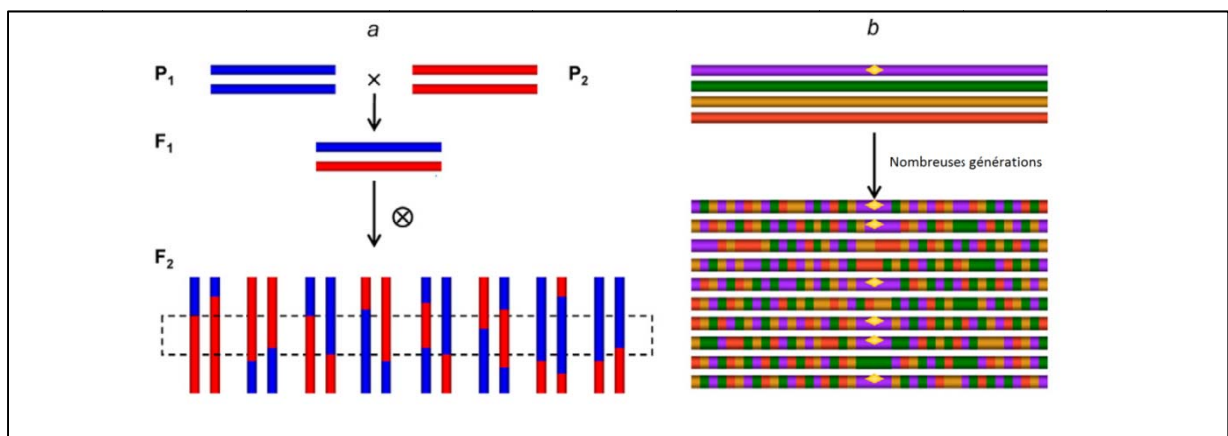


Figure 15 : Comparaison entre (a) un exercice de détection de QTL dans une population en ségrégation et (b) un exercice de détection d'associations au niveau pan-génomique. Dans le cas d'une étude d'association, l'analyse simultanée d'un grand nombre d'individus portant des allèles différents permet d'étudier l'ensemble de la variabilité génétique représentée dans la population (d'après Zhu *et al.*, 2008).

I.6.1 Avantages de la génétique d'association

La génétique d'association permet de détecter des allèles impliqués dans la variation de nombreux caractères d'intérêt à un coût relativement modeste comparativement à des études QTL, puisque les études peuvent être réalisées à partir de populations préexistantes. Il est possible, par exemple, de travailler sur des populations en sélection pour lesquelles une grande quantité de données phénotypiques est déjà collectée en routine. Selon les populations étudiées, les études d'associations donnent potentiellement accès à une diversité allélique plus ou moins grande et largement supérieure à celle étudiée dans les populations classiques de cartographie réalisées sur des croisements biparentaux (**Figure 15**). Par ailleurs, le choix de panels d'individus peu apparentés et susceptibles de renfermer un grand nombre d'évènements de recombinaison permet d'entreprendre une cartographie des allèles cibles à une échelle plus fine que les études QTL conventionnelles (Flint-Garcia *et al.*, 2003 ; Yu and Buckler, 2006).

Cependant, l'analyse simultanée de nombreux allèles contribuant au phénotype rend plus difficile l'identification des allèles en fréquence rare, en particulier ceux dont l'effet sur le phénotype est modeste (Flint-Garcia *et al.*, 2005).

L'une des principales exigences des études d'associations réside dans le choix de modèles statistiques d'analyses appropriés à la prise en compte de la structure des populations étudiées, afin de limiter le risque d'erreurs de type I (détection d'associations ne correspondant pas à des liaisons génétiques) et d'erreurs de type II (non détection d'associations réelles)

I.6.2 Paramètres pris en compte dans les études d'associations

a. Le déséquilibre de liaison

Le déséquilibre de liaison (DL), qui mesure l'association non-aléatoire entre allèles à différents loci, est un paramètre essentiel qu'il convient de mesurer avant toute étude d'association. Selon l'étendue moyenne du DL, les études peuvent s'orienter soit vers des approches « gène candidat », qui visent à associer le polymorphisme de séquences candidates à la variation phénotypique, soit vers des approches de type « genome wide » (Rafalski,

2002 ; Zhu *et al.*, 2008). La première approche est souvent privilégiée quand le DL est faible, alors que la seconde est plus aisée quand le DL est élevé. Dans le second cas c'est l'estimation de la vitesse de décroissance du DL en fonction de la distance qui détermine la densité souhaitable de marqueurs nécessaires à une bonne couverture du génome (Rafalski, 2002 ; Ersoz *et al.*, 2007).

Le DL est influencé par de très nombreux facteurs, dont le taux de recombinaisons, les événements de mutations, les phénomènes de dérive génétique dont celle provenant d'effets de fondations (Flint-Garcia *et al.*, 2003). Les espèces autogames, caractérisées par un degré d'homozygotie plus élevé que les espèces allogames, auront tendance à montrer des valeurs de DL moyen plus élevées que ces dernières en raison d'un taux de recombinaison perceptible moindre (Flint-Garcia *et al.*, 2003). Au sein d'une espèce, l'étendue du DL peut très fortement varier selon le type de population considéré. Par exemple chez le maïs, on peut trouver, dans des populations élites issues de sélection, des valeurs de DL 50 fois plus élevées que dans des populations constituées d'une large diversité de fonds génétiques (Remington *et al.*, 2001 ; Ching *et al.*, 2002). Par ailleurs, dans une population donnée, le DL peut être très hétérogène le long du génome (Flint- Garcia *et al.*, 2003 ; Ersoz *et al.*, 2007) en raison de la variabilité du taux de recombinaison et de la répartition irrégulière des zones du génome ayant subi les effets de la sélection.

Outre l'étendue du DL, la puissance d'un exercice d'associations dépend de la taille de la population étudiée, de la taille de effets des allèles recherchés, de la fréquence de ces derniers et de la qualité des données de phénotypage (Long and Langley, 1999 ; Myles *et al.*, 2009).

b. Structure et Kinship

Les études d'associations nécessitent de prendre en considération deux paramètres importants : la structure éventuelle de la population (Q) et les apparentements entre individus (K).

- La **structure de la population** (Q) correspond à l'existence de groupes d'individus génétiquement plus proches. L'existence d'une structure dans une population

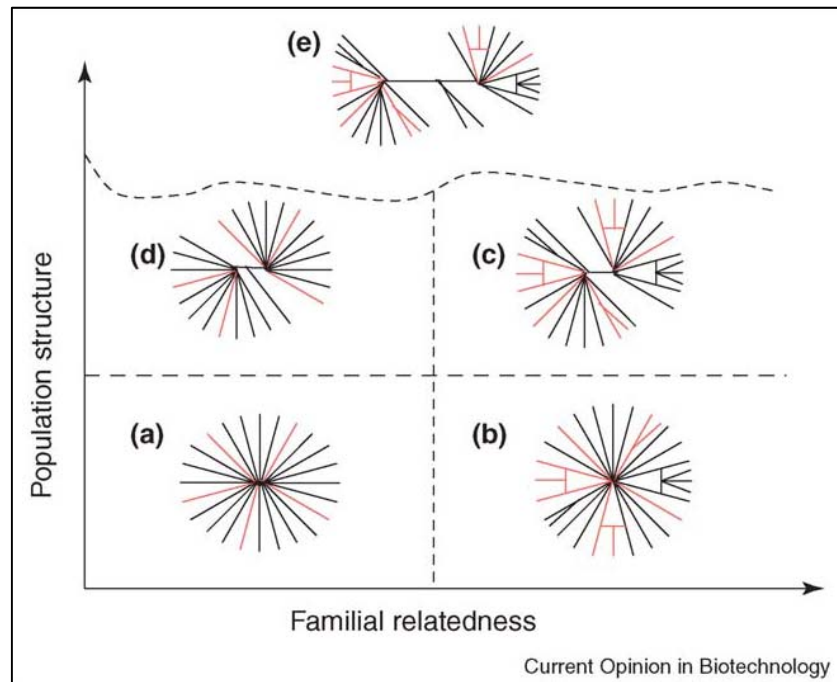


Figure 16 : Représentation schématique des différents types de populations rencontrées en génétique d'association (D'après Yu and Buckler, 2006) :

- a) Population idéale théorique, sans structure avec des niveaux d'apparentement identiques entre individus.
- b) Population non structurée dont les individus présentent des niveaux d'apparentements inégaux.
- c) Population structurée en 2 groupes avec des niveaux d'apparentements intra-groupes inégaux entre individus.
- d) Population structurée en 2 groupes avec des individus d'apparentement équivalents à l'intérieur des groupes.
- e) Population très structurée avec des niveaux d'apparentement variables entre individus.

traduit une divergence entre plusieurs groupes, résultant de la dérive génétique ou d'effets de sélection.

- Les **coefficients d'apparentement** entre individus (**K**)³, ou Kinship, traduit la structure fine dans la population relative aux liens de parenté (plus ou moins étroits) entre individus.

La **Figure 16** est une représentation schématique des notions de structure et d'apparentement, proposée par Yu and Buckler (2006).

Plusieurs méthodes statistiques, basées sur les données de génotypage des individus, permettent de détecter d'éventuelles structurations dans une population :

Pritchard *et al.* (2000a et b) ont développé un logiciel (STRUCTURE®) permettant de déterminer pour chaque individu une probabilité d'appartenance à un ou plusieurs groupes par un algorithme de clustering Bayésien. Néanmoins, une connaissance a priori du nombre de groupes probable est nécessaire pour éviter de devoir tester de trop nombreuses hypothèses de structure. De plus, cette méthode ne permet pas de tester l'hypothèse d'une absence de structure. Par ailleurs, elle est basée sur des calculs potentiellement laborieux. Cette méthode est donc peu adaptée à l'étude d'une population constituée d'un grand nombre d'individus entre lesquels il existe des relations complexes.

Patterson *et al.* (2006) et Price *et al.* (2006), ont proposé une méthode alternative basée sur une analyse en composantes principales (ACP) permettant de tester la présence d'une structuration dans une population et de quantifier son effet sur la variation du phénotype. Cette méthode a l'avantage d'être utilisable sur des populations de grande taille et issues de mélanges complexes. L'ACP est réalisée sur l'ensemble de la population, avec un sous ensemble de marqueurs non ou faiblement liés, répartis aléatoirement sur l'ensemble du génome. L'ensemble de la variabilité génétique de la population est résumé dans les axes de l'ACP. Un test de Tracy-Widom permet de déterminer le nombre d'axes de l'ACP les plus représentatifs de la structure (Patterson *et al.*, 2006). Les coordonnées des individus sur les

³ Le coefficient d'apparentement entre individus correspond à la probabilité que deux allèles tirés au hasard dans deux individus soient identiques par descendance (*c.a.d.* proviennent d'un ancêtre commun)

premiers axes (les plus significatifs) de l'ACP constituent une matrice (Q) représentative de la structure utilisable comme covariable dans les analyses d'association.

Yu *et al.* (2006) ont proposé une méthode de détection d'associations (appelée « Q+K ») basée sur un modèle linéaire mixte (MLM) dans lequel structure (Q) et Kinship (K) de la population sont utilisées comme covariables (respectivement comme effets fixe et aléatoire) pour la détection d'associations marqueurs-caractères. Ce modèle permet de réduire le nombre d'erreurs de première (α) et de deuxième espèce (β) lors d'un exercice de détection.

En outre, la multiplicité des tests d'associations réalisés dans les approches « genome scan » nécessite un second niveau de contrôle de l'erreur α , qui doit être abordée à l'échelle de l'ensemble des analyses effectuées. Ce contrôle peut être apporté par des approches basées sur des tests de permutations aléatoires des valeurs phénotypiques entre individus (Churchill and Doerge, 1994 ; Doerge and Churchill, 1996). L'autre alternative pour contrôler un risque de première espèce sur l'ensemble du génome, qui consiste à appliquer une correction de Bonferroni (Rice, 1989) [α divisé par le nombre total de marqueurs testés] possède l'inconvénient d'une perte de puissance importante (Ersoz *et al.*, 2007).

Zhao *et al.* (2007) ont montré chez *Arabidopsis*, dans le cas d'une population de structure complexe, que le modèle linéaire mixte de Yu *et al.* (2006) qui limitait au mieux le nombre de faux positifs était celui qui utilisait la méthode ACP de Patterson *et al.* (2006) pour définir les facteurs représentatifs de la structure (Q) combinée à une matrice de Kinship (K) correspondant à la proportion d'haplotypes communs entre individus.

1.6.3 Etudes d'associations chez la canne à sucre

a. Structure du DL

Il existe de nombreux indices de mesure du DL. Les plus couramment utilisés chez les espèces diploïdes (D' , r^2 , d^2) sont basés sur l'écart entre fréquences alléliques observées et attendues à l'équilibre de Hardy-Weinberg. Chez la canne à sucre, ces mesures du DL ne sont pas utilisables puisque la fréquence des allèles, et donc des haplotypes, n'est pas accessible en raison de la ploïdie élevée et de la nature dominante des marqueurs (Raboin *et al.*, 2008). La mesure du DL est basée sur des calculs de probabilité d'indépendance entre marqueurs au

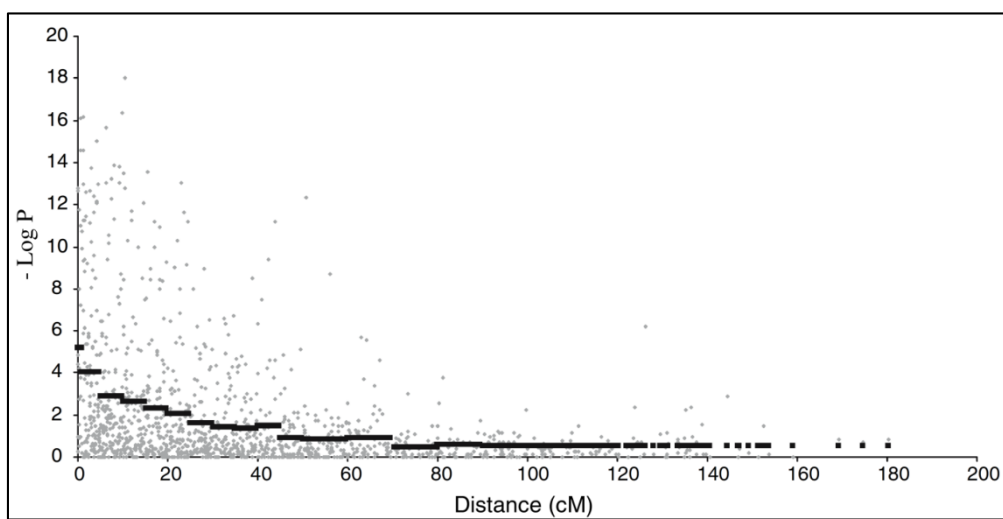


Figure 17 : Décroissance du déséquilibre de liaison (DL) le long du génome dans une population de cultivars de canne à sucre (*Saccharum* spp.) : probabilité d'association bi-marqueurs (-Log P ; test exact de Fisher) en fonction de la distance génétique entre marqueurs (en cM). Les distances entre marqueurs (AFLP) sont basées sur la carte de référence R570 (d'après Raboin *et al.*, 2008).

moyen de tests exacts de Fisher. Dans le contexte de la canne à sucre, les associations entre marqueurs ne peuvent pas toutes être détectées en raison de l'effet de confusion dû à la polyploïdie (Raboin *et al.*, 2008).

Deux études du DL réalisées sur un panel de cultivars et s'appuyant sur des données de cartographie ont été publiées chez la canne à sucre. L'étude préliminaire de Jannoo *et al.* (1999a), entreprise sur un petit panel de 59 cultivars avec un nombre limité de 38 sondes RFLP, a révélé des associations entre loci distants de quelques cM à plusieurs dizaines de cM. L'étude de Raboin *et al.* (2008), réalisée sur une base expérimentale plus étoffée (72 cultivars et 408 marqueurs AFLP cartographiés), a permis de mettre en évidence des marqueurs en DL couvrant une distance variant entre 0 et 40 cM, mais avec une décroissance rapide de la fréquence des associations significatives entre marqueurs quand la distance entre marqueurs augmente au-delà de 5 cM (**Figure 17**). Les résultats montrent une relative hétérogénéité de l'étendue du DL au sein du génome, avec quelques haplotypes de grande taille provenant d'une même origine ancestrale, issus d'une ségrégation en bloc, et de nombreux petits haplotypes correspondant à des chromosomes plus fragmentés.

Une dernière étude parue récemment qui porte sur plusieurs centaines d'accessions confirme l'intensité du DL et montre une cinétique de sa décroissance parfaitement semblable à celle qu'ont établie Raboin *et al.* (2008) (Wei *et al.*, 2010).

Ces résultats sont en accord avec l'histoire récente de l'amélioration de la canne. En effet, peu de méioses se sont produites depuis les premiers croisements interspécifiques, et peu de clones ancestraux sont impliqués dans la généalogie de ces cultivars. Peu d'évènements de recombinaison ont donc pu avoir lieu depuis les premiers croisements interspécifiques.

Ces résultats intéressants suggèrent la faisabilité d'études d'associations pan-génomiques dans une population de cultivars modernes de canne à sucre moyennant un effort de génotypage raisonnable, grâce à l'existence d'un déséquilibre de liaison appréciable.

b. Associations marqueurs-caractères

La possibilité d'études d'associations chez la canne basées sur le DL a été pressentie il y a déjà quelques années (Lu *et al.*, 1994). Quatre études concrètes de type « genome scan »

Tableau 6 : Etudes d'association réalisées chez la canne à sucre.

Taille de population	Type de population	Marquage	Caractère	Etendue du DL	Structure	Nombre d'associations	R ² individuels (%)	R ² global (%) (nombre de marqueurs associés au phénotype)	Référence
77	Clones interspécifiques fondateurs	1332 RFLP, AFLP	Résistance au charbon	-	-	64	≤ 15,9	-	Butterfield, 2007
			Résistance au borer (<i>Eldana</i>)			115	≤ 20,2	-	
74	Clones contemporains	1666 AFLP, SSR, RFLP	Résistance au charbon	0-30 cM	-	59	≤ 29	-	Raboin et <i>al.</i> , 2005
154	Clones contemporains	1209 AFLP, SSR	Résistance à <i>Pachymetra chaunorizha</i>	-	Q	31	≤ 13,2	32 (4)	Wei et <i>al.</i> , 2006
			Résistance à <i>Xanthomonas albilineans</i>			46	≤ 13	32 (4)	
			Résistance au FDV (<i>Fidji Disease Virus</i>)			25	≤ 15,1	26 (5)	
			Résistance à <i>Ustilago scitaminea</i>			50	≤ 22,6	59 (10)	
480	Clones contemporains	1531 DArT	Rendement et teneur en sucres	0-30 cM	K	42-47	-	-	Wei et <i>al.</i> , 2010

ont été publiées ces dernières années (**Tableau 6**). Elles ont toutes été effectuées sur des panels de matériel élite avec 1200 à 1500 marqueurs polymorphes. Ces études portent sur des caractères de résistance à divers bio-agresseurs (charbon, échaudure des feuilles, pachymetra root, maladie de Fidji, borer) ou sur des caractères de rendement (tonnage/ha et richesse sucrière).

Deux études (Raboin, 2005 ; Butterfield, 2007) ont été conduites sur des petits panels (77 et 74 clones respectivement) et sont fondées sur des tests d'associations directes entre le phénotype et le génotype, en l'absence de structure perceptible au sein de leur population dans une analyse de diversité Neighbour-Joining Tree (NJT) basée sur des calculs de distances. L'étude de Butterfield (2007) a été réalisée avec des marqueurs AFLP neutres et des marqueurs RFLP issus d'une expression différentielle de cDNA entre des clones inoculés au charbon (*Ustilago scitaminea*) et des clones sains. Cette étude révèle quelques marqueurs qui peuvent expliquer individuellement près de 20 % de la variation de la résistance au charbon ou de la variation de la résistance au borer *Eldana*. Dans un modèle de régression multiple, un sous-ensemble optimal de 6 marqueurs peut expliquer jusqu'à 55 % de la variation phénotypique totale de chacun des 2 caractères. Le niveau élevé de ces résultats pourrait être lié à l'éventuelle existence de gènes à effets forts présents en fréquence non négligeable dans la population étudiée. Plus certainement ces résultats sont sans doute surévalués en raison de la modestie de la taille de la population sur laquelle ils sont calculés. Plus globalement ils posent la question de la pertinence de l'usage du critère de la représentation de la diversité génétique de type NJT pour juger de l'absence de structuration, sachant que dans le contexte de la canne avec des marqueurs dominants, seule une fraction de la diversité est accessible.

Deux autres études qui émanent d'une même équipe (Wei *et al.*, 2006 ; 2010) ont été réalisées sur des panels de taille beaucoup plus importante (154 et 480 clones respectivement) et ont pris en compte l'existence d'un effet de structure dans divers modèles linéaires appropriés. Dans leur première publication (2006), les auteurs expérimentent la difficulté de l'utilisation de l'algorithme de STRUCTURE dans le contexte de la canne, pour arbitrer entre différentes structures de groupes, si l'on ne dispose pas d'informations généalogiques permettant de privilégier un choix clair face à une absence de discontinuité nette entre les différentes hypothèses de structuration. Dans leur deuxième publication (2010) les auteurs abandonnent STRUCTURE en raison de l'impossibilité d'une classification en simples groupes discrets susceptibles de rendre compte de liens complexes. L'effet de structure est

fondé sur une matrice d'apparementement entre individus. Dans les deux études (2006 et 2010) la prise en compte de la structure de la population divise en moyenne par deux le nombre d'associations détectées pour presque tous les caractères étudiés. Ces résultats illustrent l'importance bien connue de la prise en compte de co-facteurs représentant la structure dans les analyses de type « genome scan » pour limiter les risques d'inflation d'associations faussement positives. Dans leur premier article, Wei *et al.* (2006) montrent que dans un panel de 154 clones, 26 % à 59 % de la variation de la résistance peut être expliquée par la ségrégation de 4 à 11 marqueurs polymorphes. L'article le plus récent (2010) repose sur l'évaluation d'une population de grande taille (480 clones) évaluée dans un dispositif expérimental répété planté dans 3 environnements. Ces données étoffées analysées dans des modèles comparatifs intégrant différents niveaux de cofacteurs montrent qu'outre la structure, la prise en compte de corrélations spatiales, ainsi que les interactions génotype x environnements peuvent être importante pour limiter le nombre de faux-positifs potentiels.

I.7 Présentation des objectifs de la thèse

L'amélioration variétale de la canne a longtemps été basée sur des méthodes de sélection empiriques. Quelques travaux visant à identifier les bases génétiques des caractères d'intérêt en sélection, effectués sur des descendances particulières, ont permis d'identifier 3 gènes majeurs relatifs à la couleur des tiges et à la résistance à la rouille, ainsi que de nombreux QTA à effet faible, relatifs aux caractères de rendements nécessairement complexes. Ces résultats d'études QTLs, bien que prometteurs, sont basés sur l'analyse d'un nombre réduit de clones (ancestraux ou d'origine interspécifique). L'expression des allèles identifiés est liée aux fonds génétiques particuliers qui ont été étudiés, et en conséquence l'appréciation que l'on peut avoir de la valeur des allèles identifiés est toute relative. De plus ces études QTLs ne permettent pas de rendre compte de la multiplicité des loci et des allèles d'intérêts impliqués dans la variation des phénotypes observés à l'échelle d'un programme d'amélioration. La génétique d'association, basée sur l'analyse de populations en ségrégation constituées d'accessions d'origines diverses, permet d'analyser la contribution d'un grand nombre d'allèles à la variation des caractères d'intérêt. Cette approche semble d'autant plus séduisante sur le plan économique et en termes d'amélioration appliquée qu'elle permet d'étudier directement les populations couramment utilisées dans les programmes de sélection.

Cette thèse s'inscrit dans le cadre des travaux de l'équipe « Canne à sucre », au sein de l'UR « Amélioration des Plantes à Multiplication Végétative », aujourd'hui intégrée dans l'UMR Amélioration Génétique et Adaptation des Plantes (AGAP) au CIRAD. Elle a pour objectifs de :

- i) vérifier la faisabilité d'études d'associations exploitant le DL appréciable existant chez la canne à sucre dans le contexte particulier d'une plante hautement polyploïde ;
- ii) évaluer la portée et les limites d'un tel exercice de dissection génétique au regard des possibilités de marquage auxquelles on a accès pour le moment sur la canne à sucre (DArT et AFLP);
- iii) discuter l'optimisation des travaux d'analyse génétique de caractères d'intérêt en vue de perspectives appliquées aux programmes d'amélioration.

Ma thèse est fondée sur l'étude de la résistance à deux maladies d'importance économiques majeures, l'échaudure des feuilles et la maladie de la feuille jaune, qui sont des caractères dont le déterminisme dans certaines accessions est potentiellement moins complexe que celui des caractères ayant trait au rendement agronomique ou sucrier. L'étude de la résistance à la première maladie est fondée sur un large panel d'association expérimenté dans un seul contexte épidémiologique (Guadeloupe). L'étude de la résistance à la seconde maladie s'appuie sur deux panels d'association distincts expérimentés dans deux environnements épidémiologiques différents (Guadeloupe et Réunion).

Deuxième partie : Etude
d'associations relative à la résistance de la
canne à sucre à la contamination aérienne
par *Xanthomonas albilineans*

I.1 Introduction

L'échaudure des feuilles est une maladie bactérienne importante de la canne à sucre, dont la sélection est réputée difficile dans la plupart des pays en raison du caractère souvent erratique de l'expression des symptômes foliaires provenant d'une infection systémique de la tige. Dans le contexte particulier de la Guadeloupe, on s'est aperçu récemment de deux éléments favorables à la propagation naturelle de la bactérie par la voie aérienne : i) la présence exclusive de populations bactériennes correspondant au génotype (B) décelé dans la plupart des pays ayant connus des épidémies récentes (Davis *et al.*, 1997 ; Champoiseau *et al.*, 2006a) ; ii) un climat favorable présentant des niveaux d'humidité permettant le déclenchement périodique d'épidémies (Daugrois *et al.*, 2003 ; Champoiseau *et al.*, 2009).

Dans la présente publication, nous avons exploité les conditions environnementales favorables de Guadeloupe à la propagation de *Xanthomonas albilineans* par voie aérienne, pour réaliser une étude d'associations relative à la résistance de la canne à la contamination aérienne de l'échaudure des feuilles sur la base de l'étude d'un panel de 189 cultivars génotypés avec 4189 marqueurs polymorphes (AFLP et DArT). L'influence des variations climatiques interannuelles sur l'appréciation des niveaux de résistance a motivé l'expérimentation du panel d'association sur plusieurs années afin de tester la reproductibilité des résultats d'analyses génétiques obtenus.

I.2 Publication associée

Rocher S., Daugrois J.H., Nibouche S., Boisé-Noc R., Hervouet C., Roques D., Toubi L., Costet L., D'Hont A., Hoarau J.Y.

Genome-wide association study for sugarcane resistance to aerial contamination by *Xanthomonas albilineans*.

Soumis à la revue *Molecular Breeding* le 22/09/2011

Genome-wide association study for sugarcane resistance to aerial contamination by *Xanthomonas albilineans*

Solen Rocher^{1,4}, Jean-Heinrich Daugrois², Samuel Nibouche³, Rosiane Boisé-Noc¹, Catherine Hervouet⁴, Danièle Roques¹, Lyonel Toubi¹, Laurent Costet³, Angélique D'Hont⁴, Jean-Yves Hoarau¹

(1) CIRAD, UMR AGAP, F-97170 Petit-Bourg, Guadeloupe, France

(2) CIRAD, UMR BGPI, F-97170 Petit-Bourg, Guadeloupe, France

(3) CIRAD, UMR PVBMT, F-97455 Saint-Pierre, La Réunion, France

(4) CIRAD, UMR AGAP, F-34398, Montpellier, France

Corresponding author: Jean-Yves Hoarau, Tel : Email: jean-yves.hoarau@cirad.fr,

Telephone: (059)05 90 94 46 02 Fax: (059) 05 90 94 88 99

Abstract

Leaf scald (caused by *Xanthomonas albilineans*) is a damageable bacterial disease for sugarcane (*Saccharum* spp.) and thus a major concern in breeding programs. We undertook a genome-wide association study to tag alleles associated with sugarcane resistance to aerial contamination by *X. albilineans*, using a panel of 187 interspecific sugarcane cultivars. This panel was assessed for Disease Severity (*DS*) in two trials (T1 and T2) carried out in Guadeloupe, on the basis of observed foliar symptoms in two successive ratoon crops under natural infection conditions. The cultivar panel was genotyped with 4189 polymorphic AFLP and DArT markers. Population structure was analyzed by a Principal Component Analysis using a subset of markers not significantly linked. Marker-trait associations were tested in mixed linear models taking into account population structure with or without kinship

information. Considering both trials environments, a total of 23 independent genomic regions were significantly associated with *DS* trait at a Genome Wide type-I Error Rate (GWER) of 0.05. Among them, six were simultaneously significant in T1 and T2 trials. However, consistency of associations between trials was good when considering the large overlapping between the best parsimonious subsets of markers revealed by stepwise multiple regressions. Each region accounted individually from 4% to 8% of total phenotypic *DS*-trait variation. Altogether 19 of the 23 detected regions accounted for 41% and 47% of leaf-scald *DS* variation in T1 and T2 trials, respectively. These regions possess genes potentially useful for leaf-scald resistance breeding programs.

Keywords : leaf scald, *Saccharum* spp., Quantitative Trait Allele, multiple regression

Introduction

Leaf scald of sugarcane (*Saccharum* spp.) is caused by *Xanthomonas albilineans*, a bacterial vascular pathogen. This major disease of sugarcane is present in most of the sugarcane producing countries in the world (Rott and Davis, 2000) and more or less severe yield losses may occur when susceptible cultivars are grown (Rott *et al.* 1995; Holder *et al.* 2000). Most recent leaf scald outbreaks reported between 1992 and 2001 were related to *X. albilineans* strains belonging to the same genetic group (Group B) that has a better propensity to epiphytic life and aerial propagation (Davis *et al.* 1997; Champoiseau *et al.* 2006).

Breeding for leaf scald resistance has long been a major concern in breeding programs (Walker 1987), since cultivation of resistant varieties is the only efficient mean to control the disease in the absence of any sustainable curative treatment of susceptible varieties. Resistance to leaf scald is an essential prerequisite for varieties to be released for commercial cultivation. Earliest possible elimination of susceptible material is applied in sugarcane breeding programs and shift toward parent varieties that give higher proportions of resistant progeny is pursued (Bechet *et al.* 1992). Traditionally, screening of material only relied on the use of artificial inoculation of field experiments (Koike *et al.* 1965). Such experiments that are labor intensive, are only restricted to the assessment of a limited number of clones at advanced selection stages. These constraints in managing artificial screening preclude the evaluation of large and numerous progenies at early stage of selection. This largely explains the absence in literature of any heritability estimates of resistance to leaf

scald and the scarcity of Quantitative Trait Allele (QTA) mapping studies. Only two QTA studies for leaf scald resistance have been conducted thanks to inoculated experiments: the first one (Offmann 2000) aiming at analyzing the determinism of resistance in an elite cultivar relied on an inoculated trial of a large ad-hoc progeny; and the second one (Wei *et al.* 2006) consisting of an association mapping study on elite genotypes using the disease susceptibility data available from a breeding program.

However, new considerations provides ways to get round constraints linked to disease inoculation : i/ the recent discovering of aerial transmission of the pathogen and contamination of plant canopy resulting in particular necrotic leaf symptoms prior to stalk infection (Autrey *et al.* 1995, Comstock *et al.* 2001, Daugrois *et al.* 2003); ii/ Guadeloupe environment features, including predominance of a *X. albilineans* genetic group B (Champoiseau *et al.* 2006) and humid climate conditions that favor efficient natural disease infection (Daugrois *et al.* 2003; Champoiseau *et al.* 2009). These features give new opportunities to easily screen relatively large amount of material needed for genetic studies. Taking advantage of natural epidemic events, screening could be based on the observation of the expression of the leaf necrotic stripe symptoms.

Molecular genetics based on genomic markers offers efficient approaches to better grasp the genetic bases underlying agronomic traits of interest, such as disease resistances. Reasoning of genetic analysis on sugarcane needs to take into account the complex, highly polyploidy and aneuploidy of plant genome (Hoarau *et al.* 2007). Modern sugarcane cultivars (*Saccharum* spp.) are the results of complex interspecific hybridizations between *Saccharum officinarum* ($x=10$, $2n=8x=80$ chromosomes), the sugar-producing species, and the wild *Saccharum spontaneum* species ($x=8$, $2n=5x-16x=40$ to 128 chromosomes) (D'Hont *et al.* 1996; 1998). Due to a $2n + n$ chromosome transmission occurring in crosses and first backcrosses between these two ancestral species, modern sugarcane genome encompass between 100 and 120 chromosomes, and about 15-27% of it is inherited from the wild species (Piperidis *et al.* 2010). Each of the 10 homo(eo)logy groups has around 9 to 14 chromosomes (Grivet and Arruda 2001).

Despite this large and complex polyploid genome, numerous QTL studies have been conducted using progenies for several agronomic traits (Hoarau *et al.* 2002; Reffay *et al.* 2005; Piperidis *et al.* 2008; Aitken *et al.* 2008; Nibouche *et al.* in press), thanks to relatively frequent available single-dose markers related to the high heterozygosity of varieties (Lu *et al.*

1994). Moreover, recent studies have shown the existence of important linkage disequilibrium (LD) among modern cultivars (Jannoo *et al.* 1999; Raboin *et al.* 2008) in agreement with the recent breeding history of the plant and its limited number of founders. LD, which can span up to 10 cM or more in a few instances, decreases sharply after 5 cM on average. This situation makes feasible LD-based association studies owing a reasonable effort of genotyping (Wei *et al.* 2006; Wei *et al.* 2010). Compared to bi-parental QTL approaches, association mapping offers the advantage of exploring allelic diversity on a large number of accessions.

In the present study, we took advantage of the favorable ecological conditions to epiphytic spread of leaf scald pathogen prevailing in Guadeloupe, to undertake a genome-wide association mapping study on a panel of modern cultivars, in order to tag useful markers involved in the genetic determinism of sugarcane resistance to aerial contamination by *X. albilineans*.

Materials and methods

Cultivar panel and field experimentations

A panel of 189 interspecific sugarcane cultivars (*Saccharum* spp.), representative of the germplasm used in breeding programs worldwide, was selected for the present study. This material, obtained from several sugarcane breeding centers, is listed in Sup. Tab. 1. The cultivars were chosen on the basis of their limited genealogical relationships, thanks to available genealogical information (Machado 2000; Ruiz *et al.* 2004, <http://www.tropgenedb.cirad.fr/>).

The 189 cultivar panel was experimented in Guadeloupe at Cirad Roujol Station, in two trials (T1 and T2), each one consisting of a complete randomized block design with 3 replications. Basic plot was a 2.5 m row and rows were spaced 1.5 m apart. T1 and T2 trials were planted in September 2005 and September 2007 respectively and were cultivated during three successive crop-cycles following commercial agronomic practices. Random arrangements of the cultivars within replications were different between both trials. Both trials were harvested each year before the rainy season (in May 2006 and March 2007 for T1 and in June 2008 and May 2009 for T2). The cultivar panel was phenotyped for disease foliar symptoms once each year, in first ratoon [R1] and second ratoon [R2] crops in both trials,

after natural aerial contamination at the end of the raining season (October/November), when maximum symptoms intensity is expected. The two successive trials observed during two ratoon crops, represented four annual epidemiological seasons (2006, 2007, 2008, 2009). Disease was assessed for each clone by measuring symptoms on all leaves of 10 randomly chosen stalks in each three replications using a 0-4 scoring scale, where 0 = no symptom, 1 = one or two small necrotic lesions (<10cm in length), 2 = more than two small necrotic lesions (<10cm in length) or one necrotic lesion between 10 and 30 cm in length, 3 = several necrotic lesions between 10 and 30 cm in length or one lesion > 30 cm in length, 4 = more than one lesion > 30 cm in length. Disease Severity (DS) of each basic plot was expressed in percentage as $DS = (1 \times N_1 + 2 \times N_2 + 3 \times N_3 + 4 \times N_4) / 4 \times N_T$, where N_i = number of stalk with score i and N_T = total number of stalks. So DS ranged from 0% (no symptoms) to 100% (all stalks scored 4).

Genotyping

The cultivar panel was genotyped with a total of 4189 polymorphic markers among which 1713 DArT markers and 2475 AFLP markers. We performed AFLP genotyping using the AFLP® Analysis System I (Invitrogen) by the use of 59 primer pairs as recommended by the manufacturer with slight modifications described by Hoarau *et al.* (2001). DArT whole genome profiling of the cultivar panel was produced by the commercial company Diversity Arrays Technology Pty Ltd (Yarralumla, Australia) using their commercial sugarcane array (Heller-Uszynska *et al.* 2011).

Statistic analysis

Assessment of the genetic value of the cultivar panel for leaf-scald resistance

Analyses of Disease Severity (DS) recorded in R1 and R2 ratoon crops after annual canopy natural infection were performed separately for each trial, at both crop-year level and across ratoon crops with the two following mixed linear models:

$$DS_{ij} = \mu + G_i + R_j + E_{ij} \text{ (model 1)}$$

$$DS_{ijk} = \mu + G_i + R_{jk} + CY_k + G_i \times CY_k + E_{ijk} \text{ (model 2)}$$

where:

DS_{ij} (DS_{ijk}), is the observation of genotype i in replicate j (in a particular crop-year k), μ is the overall mean, G_i is the random effect of genotype i , with variance σ_G^2 , CY_k is the fixed effect of crop-year k , R_j (R_{jk}) is the fixed effect of replication j (in a particular crop-year k), $G \times CY$ is the random interaction effect between genotype i and crop-year k , with variance $\sigma_{G \times CY}^2$, E_{ij} and E_{ijk} are residuals of respective models, with variance σ_e^2 .

To achieve normal distributions of residuals, DS percentage values were transformed using arc-sinus root-square transformation prior to analyses. Analyses were performed with SAS procedure Mixed (SAS Institute 2008). Significance of random effects was assessed with a z-test (covtest option in proc Mixed). In the second model, correlations of residuals between years due to observations carried out on the same basic plots (longitudinal data) was taken into account with the use of the repeated statement of the Mixed procedure. BLUPs (Best Linear Unbiased Predictors) were then estimated by adding the overall mean μ to the solution of the random cultivar effect of each model and these values were used to perform marker trait association analyses.

In order to assess the accuracy of the BLUP values given by the experimental design herein used, we estimated the broad-sense heritability (H^2) at an experimental level, using variance component estimates attached to model 1 [$H^2 = \sigma_G^2 / (\sigma_G^2 + \sigma_e^2 / 3)$] and model 2 [$H^2 = \sigma_G^2 / (\sigma_G^2 + \sigma_{G \times CY}^2 / 2 + \sigma_e^2 / 6)$]. Single plot-based broad-sense heritability (h^2) values were also calculated [$h^2 = \sigma_G^2 / (\sigma_G^2 + \sigma_e^2)$] for comparison.

Panel genetic structure

Genetic structure of the cultivar panel was inferred from a sub-set of DArT markers, using the Principal Component Analysis (PCA) approach developed by Patterson *et al.* (2006). Following this approach, the genome-wide markers subset considered for the PCA was restricted to the only markers not found in tight LD between each other. Using Fisher Exact tests to test associations between markers (freq procedure of SAS Institute 2008), and considering the Bonferroni correction for multiple testing (Rice 1989), we defined a subset of 867 markers not significantly linked. Markers scoring data were standardized and missing data were replaced by zero value before computing PCA with the Princomp procedure of SAS

software (SAS Institute 2008). Two cultivars were considered as outliers in the PCA analysis, i.e. their coordinates in at least one of the ten first axes were five times greater than the within axe standard deviation of the whole cultivar panel coordinates. Both cultivars were therefore removed from the analysis leading in a final panel of 187 cultivars. The number of significant PCA axes representative of genetic structure was defined by a Tracy-Widom test at a 0.05 error level (Patterson *et al.* 2006).

Marker-trait associations

Detection of marker-trait associations were carried for each trial considering BLUP values estimated across both ratoon crops by model 2, which took into account the repeated nature of data between years (same individual plot location within replicates). Unified mixed model approaches of Yu *et al.* (2006) were applied using TASSEL software 2.1 (Bradbury *et al.* 2007). Analyses were performed with the GLM model (using coordinates along PCA axes as fixed covariables) and with the MLM model (same as GLM plus a random polygenic effect added through a variance-covariance matrix derived from a kinship matrix computed in TASSEL).

The number of PCA axes considered as fixed covariables in both GLM and MLM models were chosen following Zhu and Yu (2009) method among the first significant PCA axes contributive to genetic structure previously identified by a Tracy-Widom test (Patterson *et al.* 2006). Using a linear model, each BLUP set was fitted with an increasing number of PCA axe coordinates, and the analysis showing the lowest Bayesian Information Criteria (BIC) determined the best number of dimensions to be considered in mixed models.

Finally, significance of marker-trait associations were considered using (i) nominal *P* value, (ii) *P* value at a genome wise type-1 error rate (GWER) of 0.05 (based on 2000 permutations) and (iii) consistency of significant associations between trials.

Significant markers were assessed for pair-wise marker associations using Fisher exact tests and considering the Bonferoni procedure (Rice 1989) for multiple tests corresponding to a global type-I error of 0.05. Markers that were associated at this error level were considered belonging to a same haplotype following Raboin *et al.* (2008) approach.

Proportion of disease resistance variance explained by markers

In order to estimate the global effect of the marker sets individually associated to leaf-scald resistance trait, we achieved for each trial a multiple regression of *DS* genotype values (BLUPs) on all independent markers detected at a GWER of 0.05 (selection of the most significant marker in each haplotype) and having less than 5% missing molecular data. Moreover, the best parsimonious subset of markers altogether accounting for the highest contribution to the total phenotypic variance was identified by using Stepwise option of SAS Reg procedure. This procedure added (or deleted) markers one by one to the model, when the p-value of the F statistic reflecting contribution of the marker to the model, was lower (or greater) than a threshold value fixed at 0.15.

Results

Leaf-scald (LS) assessment of the cultivar panel

The present genetic analysis of leaf-scald (LS) resistance determinism was based on the study of a panel of 187 cultivars phenotyped in two trials (T1, T2) in first ratoon (R1) and second ratoon (R2) crop-year cycles following natural local disease infection after the rainy season. Analyses of variance (ANOVA) were achieved within trials at crop-year (model 1) or across crop-year (model 2) levels. Considering all ANOVAs, genotype effects were always highly significant ($P < 0.0001$) for LS Disease Severity (*DS*) trait measuring foliar infection (table 1). Within year distribution of cultivar *DS* means showed a general “L shape” distribution (Figure 1) less pronounced in 2006 (T1-R1) than in the three other years [2007 (T1-R2), 2008 (T2-R1), and 2009 (T2-R2)]. Corresponding average *DS* of the panel was about two-fold higher in 2006 (T1-R1) compared to the three other years (2007, 2008 and 2009) which showed similar average levels. Congruently, in the second ANOVA model, crop-year effect was highly significant ($P < 0.0001$) in T1 but not in T2 (table 1). Broad-sense heritability of *DS* at an individual plot level was comparatively higher (0.75) in 2006 (T1-R1) compared to the three following years (0.60-0.68 range). Concerning broad-sense heritability values at an experimental level, this parameter which is based on an entry-mean basis of three replicates ranged in high values in all four years (0.82-0.90).

ANOVA across crop-cycles corresponding to model 2, showed a highly significant ($P < 0.0001$) interaction between genotype (*G*) and crop-year (*CY*) effect in each trial.

However, this *GxCY* interaction effect appeared very relative since i) broad-sense heritability of *DS* trait values (0.89 for T1 and 0.83 for T2) at an experimental level across years (model 2) were as high as heritability values within year (model 1) and ii) phenotypic correlations of *DS* trait between first (R1) and second (R2) ratoons in both trials were also relatively high [0.78 ($P<0.0001$) and 0.64 ($P<0.0001$) in T1 and T2, respectively].

Marker-trait associations within trials

The analysis of the structure of the cultivar panel based on a PCA, revealed 15 significant axes ($P<0.05$) using the Tracy-Widom statistic. Fitting the BLUP of *DS* values of each trial (obtained from model 2) in multilinear regression models with an increasing number of PCA axes, led to a minimum BIC value corresponding to 4 and 7 axes in T1 and T2 trials respectively. The first 4 axes explained $R^2=11.34\%$ ($P<0.0002$) of *DS* variation in T1, while the first 7 axes explained $R^2=15.61\%$ ($P<0.0001$) of *DS* variation in T2. These numbers of PCA axes were further used as covariates in GLM and MLM models to correct for the structure of the panel.

Table 2 summarizes the marker trait detections significant at a Genome Wide Error Rate (GWER) of 0.05 in each trial using GLM model. Considering both trials together, a total set of 29 markers appeared significantly associated with *DS* trait at this GWER. Among them 17 were significant in T1 trial while 16 were significant in T2. Corresponding P nominal values of the considered GLM model ranged from 7.02×10^{-5} to 3.8×10^{-3} . P nominal values obtained with the MLM model, were on average 48% higher compared to values revealed with GLM model. The 29 detected markers represented five haplotypes (grouping two or three markers) and 18 independent markers corresponding to a total of 23 independent chromosomal regions of agronomic interest within the surveyed panel genomes. Among these 23 regions, 14 and 13 were detected significantly ($P_{\text{GWER}}<0.05$) in T1 and T2, respectively. Among this 27 total region-trait associations, 22% (6 over 27) were significant at $0.001<P_{\text{GWER}}<0.01$ and another 37% (10 over 27) were significant at $P_{\text{GWER}}<0.001$.

Consistency of genetic effect across trials

The detection of genetic effects significant at a P_{GWER} of 0.05 simultaneously in both trials appeared modest, since among the 14 or 13 significant regions detected at this threshold in T1 and T2 respectively, only four were common to the two trials. The stability of associations was further investigated by decreasing the stringency and allowing associations

with lower significance to appear. This exploratory investigation was successful, since when decreasing the GWER threshold up to 0.10, we recovered four additional common associations between trials (table 2) and three other additional ones at $P_{\text{GWER}} < 0.15$ (data not shown). Among the 23 detected chromosomal regions, 10 showed a favorable effect on leaf scald resistance and 13 an unfavorable effect. According to the relatively high phenotypic correlation of *DS* values between both trials ($R=0.83$; $P<0.0001$), the associations were always in the same direction in both trials.

Leaf-scald resistance variation explained by markers

The proportion of the total phenotypic variation (R^2) explained individually by the 23 detected regions was generally small and ranged from 4% to 8%, except for one marker (acactc24) detected in T2, whose R^2 of 13% is obviously over-estimated due to a large number of missing genotypic data. In order to avoid any over-estimation of the global marker effect, multiple linear regressions were computed with the only 19 markers from the 23 independent regions that had less than 5% of missing coding data (four markers with less than 179 cultivars genotyped discarded). As shown in table 3, this set of markers explained altogether 41% ($df=133$) of the total phenotypic variance in T1 and 47% ($df=130$) in T2 trial. The use of a stepwise multiple regression approach permitted to identify a best parsimonious subset of nine and seven markers in the two respective trials that explained altogether 36% ($df=155$)[T1] and 40% ($df=160$)[T2] of the total phenotypic variance. These two regressions shared six common markers (actcaal4, agccat11, B128918, H367877, H419516 and H427561).

Discussion

The present association study aimed at tagging genomic regions involved in resistance to the aerial contamination of sugarcane by leaf scald pathogen. To this end, we studied a large panel of cultivars representative of modern sugarcane cultivars used in breeding programs worldwide. This panel was tested thanks to natural inoculations in the favorable epidemiological conditions of Guadeloupe.

Because of the well-known climatic inter-annual variations prevailing locally, we chose to phenotype the studied cultivar panel repeatedly over times, to assess the consistency

of the symptoms across years and the resulting influence on the allele tagging exercise. The two successive trials observed during two ratoon crops represent four annual epidemiological seasons (2006, 2007, 2008 and 2009). A higher disease pressure level was observed in 2006 compared to the three following years (2007, 2008 and 2009), and can explain the higher broad-sense heritability of *DS* observed at a plot level in 2006 (0.75) compared to the three other years (0.60-0.68 range). However, the main finding to be emphasized resides in the high broad-sense heritability of *DS* observed at an experimental level in each of the four years studied (0.82-0.90) as well as at trial level across both ratoon crops (0.89 in T1, 0.83 in T2). This indicates that the three-replicate experimental design herein used was efficient to control environmental errors, and to compute accurate BLUP genetic estimates of the resistance level of the surveyed cultivars. These results clearly indicate that in appropriate ecological environments, dispersion of the pathogen within trials due to aerial leaf scald contamination is homogeneous. These results are consistent with recent epidemiological studies (Champoiseau *et al.* 2009) that showed that canopy can be colonized within few weeks by *X. albilineans*.

Relatively high phenotypic correlations of *DS* values were observed between ratoon crops within trials (0.78 in T1 and 0.64 in T2). These findings indicate that the variation of the inter-annual epidemiological conditions that prevailed in each trial had a modest impact on the repeatability of disease indexation, providing the use of appropriate replicated trials.

The resistance/susceptible allele tagging exercise in each trial was based on the use of more than 4000 neutral dominant markers (DArT and AFLP). The use of an analysis model based on the phenotypic data across ratoons within both trials provided two Quantitative Trait Allele (QTA) detections opportunities. The absence of a significant change of scale in the magnitude of *P* nominal values between the two models used (GLM, taking into account population structure; MLM, taking into account both population structure and kinship information) indicates a modest effect of the genetic relationships between cultivars on the marker-trait-detections. This finding could be related with the fact that the panel was chosen on the basis of limited genealogical relationships between cultivars. Considering the two tagging exercises, a total of 23 haplotypes or independent marker alleles were found associated with the disease trait at a relatively conservative genome-wise type-I error rate (P_{GWER}) of 0.05, which permitted to tag the most probable genuine regions of agronomic interest. The proportion of regions which were commonly detected at this threshold between the two trials was less than a third (4 over 14 or 13) but reached about 80% (11 over 14 or 13) when relaxing the P_{GWER} threshold up to 0.15 for an explanatory purpose. These results

illustrated the fact that in the context of panel association study, where quantitative effects of alleles of interest are all expected to be small as the panel size is large (particularly in the highly polyploidy context of sugarcane), allele detection appears very sensitive to the statistical threshold that truncates the allele repertoire by retaining the only highest effects. However, despite a frequent highly fluctuation of both significance and effect size of markers between the two trial environments, the consistency of the general QTA picture is good regarding the high proportion of overlapping markers between the two best marker subsets.

In our association mapping study, the size of individual marker effects (R^2) ranged between 4% and 8%. Interestingly, this range of R^2 is exactly identical to the one of the QTAs detected by Offmann (2000) in cultivar 'R570' based on simplex AFLP markers analyzed on a self progeny of this cultivar (150 individuals). Considering the similar population size contexts we would have expected smaller QTA in the association study due to a much larger number of alleles segregating in the panel at the loci governing the trait. Two reasons, possibly complementary, might explain the equal range of R^2 in both QTA studies: (i) genes with strong effects might exist in some cultivars of the panel but be absent in R570 (ii) in the association study most of the polymorphic dominant markers considered for associations have various dosage in the cultivars that bear them. Therefore R^2 of most markers represent the effect of more than one allele copy, where as in the QTL study R^2 correspond to the effect of a single-dose allele.

Our mapping association study revealed that the only segregation of a small subset of 9 to 7 independent markers could account for 36% to 40% of the variation of leaf scald resistance observed in the studied panel. Besides, half of these markers have a rather low individual frequency (0.09-0.18 range). Among those, some might be in single dose in most of the accessions that bear it, and could tag QTA that might correspond to alleles of major effects in bi-parental derived progenies. Therefore, our results offer key information to target candidate crosses, for the identification of possible major resistance QTAs, useful to optimize breeding strategies for leaf-scald resistance.

Our association mapping results could be interestingly compared with those of Wei *et al.* (2006), who also tagged resistance alleles to sugarcane leaf-scald by studying a panel of 154 accessions genotyped with about 1200 polymorphic AFLP or SRR markers. Differences exist between their work and our present ones, concerning nature of the panel (mainly Australian elite material vs. international cultivars), phenotypic data origin (from numerous

trials compared to two trials), mode of disease infection (artificial vs. natural inoculation), statistical models used to test marker-trait associations, and approaches used to take into account population structure [cofactors from STRUCTURE program (Pritchard *et al.* 2000) vs. from PCA analysis]. However, when considering population structure, Wei and al. (2006) finally identified six markers at a nominal $P \leq 0.001$ threshold. Five of these markers explained altogether 26.2% of the total phenotypic variance in a multiple regression based on 123 total degree of freedom (*df*). Comparatively in our study, we identified 19 independent markers at a GWER threshold (0.05) corresponding to P nominal values (7.02×10^{-5} – 3.8×10^{-3}) equivalent to threshold of Wei *et al.* (2006)'s paper ($P \leq 0.001$). The global effect of these 19 markers explained 41% (T1) and 47% (T2) of the disease variation in multiple regressions based on identical panel size ($df=123$ compared to 133 or 131). The explained variance that is about 1.5 fold higher in our work is probably the result of a genotyping effort about 3.5 fold larger (4189 vs. 1200 markers). This indicates that an improvement of the dissection of leaf-scald resistance determinism is still feasible owing further investments in genotyping efforts.

Considering the extensive linkage disequilibrium (LD) existing in modern germplasm (Raboin *et al.* 2008) that facilitates association mapping, the scale of the effort needed to further improve the present genetic dissection, would mainly depend on the nature (quantitative vs. qualitative) of the factors underlying the remaining untagged fraction of the variance. Besides further intensification of the genome-wide association approach, it would be profitable to develop a strategy based on targeted associations using coming information about *Xanthomonas*/grasses pathosystems (Gu *et al.* 2005) and syntenic relationships between sugarcane and other more advanced model grasses.

Acknowledgements

The authors thank Jean-Marie Coupan and his team for technical field assistance. This work was funded by Centre de Coopération Internationale en Recherche Agronomique pour le Développement, by the Conseil Régional de la Guadeloupe and by European Union (FEDER – DRRT fund).

Figure and tables

Table 1 : Summary statistics for leaf-scald Disease Severity (*DS*) trait in both trials (T1, T2) in the two ratoon crop-years (R1, R2) from analyses of variance: mean value, range and standard deviation within the cultivar panel, broad-sense heritability of *DS*, genotypic effect (*G*), crop-year interaction (*CY*) effect and (*G*×*CY*) interactions.

Items	Trials					
	T1			T2		
	R1	R2	Across	R1	R2	Across
	(2006)	(2007)	R1/R2	(2008)	(2009)	R1/R2
	Model 1		Model 2	Model 1		Model 2
<i>DS</i> mean of the cultivar panel	0.26	0.14	0.19	0.12	0.11	0.11
<i>DS</i> range	0-0.91	0-0.77	0-0.83	0-0.81	0-0.70	0-0.63
<i>DS</i> standard deviation	0.21	0.16	0.17	0.17	0.13	0.13
Genotypic (<i>G</i>) effect ^a	***	***	***	***	***	***
Crop-year (<i>CY</i>) effect ^a	-	-	***	-	-	NS
<i>G</i> × <i>CY</i> interaction ^a	-	-	***	-	-	***
Broad-sense heritability (h^2) at a single plot level	0.75	0.68	0.69	0.68	0.60	0.59
Broad-sense heritability (H^2) at an experimental level	0.90	0.87	0.89	0.86	0.82	0.83

(a) : ***: $P < 0.0001$; NS: $P > 0.05$

Figure 1: Distribution of leaf-scald Disease Severity (*DS*) among cultivar panels in both trials (T1, T2) and in both ratoon crop-years (R1, R2).

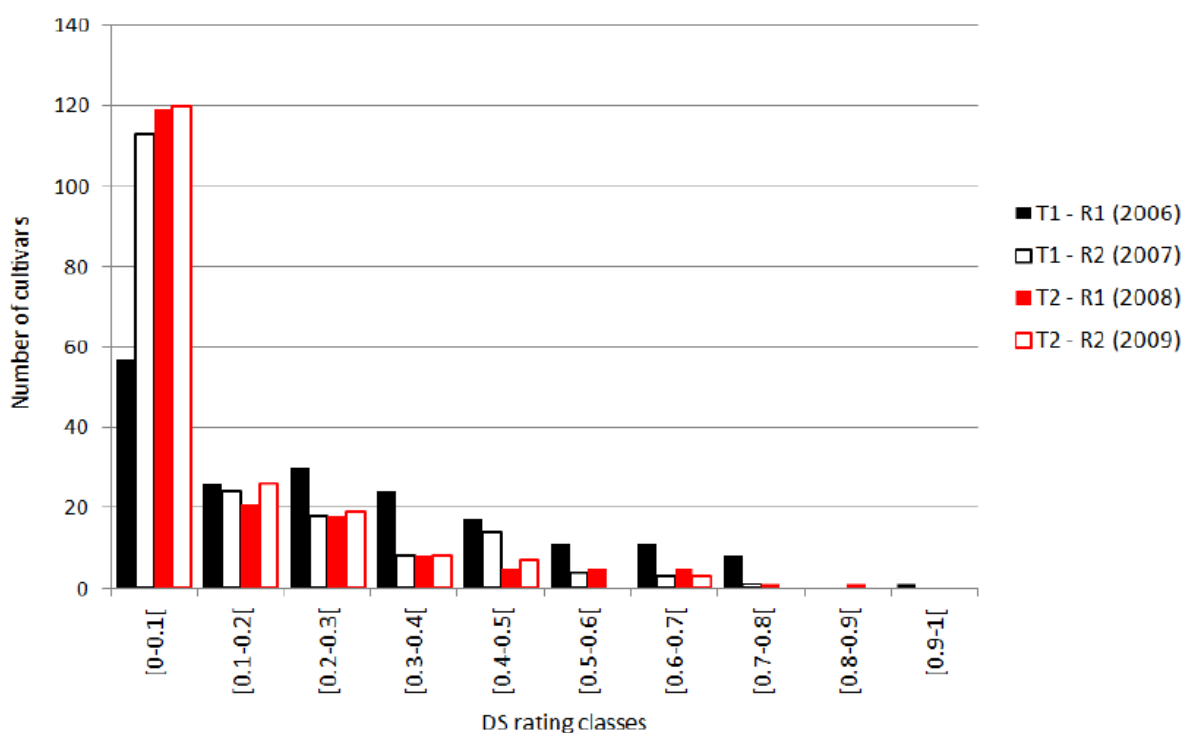


Table 2 : List of markers significantly associated with leaf-scald resistance at a GWER of 0.05 in at least one trial (T1, T2) across the two ratoon crop-years (R1, R2), using correction for structure (GLM model).

Marker ^a	Marker frequency	Number of genotyped cultivars	Trial 1 (T1)				Trial 2 (T2)				Effect ^d	Markers simultaneously significant in both trials at a GWER of 0.05	Markers significant in one trial at a GWER of 0.05 and in the other one at a GWER of 0.10
			Across R1/R2			MLM ^c	Across R1/R2			MLM ^c			
			GLM				GLM						
			Nominal P value	P-permut ^b	R ²		Nominal P value	P-permut ^b	R ²				
H420097	0.10	182	3.67E-04	***	0.06	3.70E-04	3.40E-03	NS	0.04	2.60E-03	S		x
B420097	0.10	185	1.40E-03	*	0.05	1.10E-03	1.31E-02	NS	0.03	6.80E-03	S		
Aggcaa33	0.37	179	7.02E-05	***	0.08	7.40E-05	1.70E-03	*	0.05	1.70E-03	S	x	
Aagctt28	0.15	187	2.80E-03	*	0.04	5.40E-03	1.21E-02	NS	0.03	1.78E-02	S		
H382397	0.50	185	2.20E-03	*	0.05	1.40E-03	2.97E-02	NS	0.02	2.21E-02	R		
Acccat2	0.40	177	9.67E-04	**	0.05	1.50E-03	5.60E-03	NS	0.04	1.03E-02	R		
H257517	0.16	187	1.28E-02	NS	0.03	1.60E-02	7.54E-04	**	0.05	1.40E-03	S		
H370204	0.16	186	2.57E-02	NS	0.02	3.40E-02	1.70E-03	*	0.05	3.60E-03	S		
H367877	0.16	185	8.00E-03	NS	0.03	1.18E-02	4.27E-04	***	0.06	1.60E-03	S		
H416781	0.18	182	9.00E-03	NS	0.03	1.43E-02	6.50E-04	**	0.06	1.30E-03	S		
H370244	0.21	177	7.78E-02	NS	0.02	1.03E-01	1.90E-03	*	0.05	3.10E-03	S		
H427561	0.09	186	2.44E-02	NS	0.02	4.74E-02	1.50E-03	*	0.05	3.60E-03	S		
accac54	0.25	186	3.80E-03	NS	0.04	5.10E-03	1.70E-03	*	0.05	2.50E-03	S		x
aggcac11	0.61	167	7.42E-04	**	0.06	6.87E-04	4.71E-04	***	0.06	1.00E-03	R	x	
aggett1	0.80	131	2.20E-03	*	0.06	2.80E-03	5.21E-04	**	0.08	5.37E-04	R	x	
H419516	0.10	186	4.31E-04	***	0.06	7.03E-04	2.90E-03	NS	0.04	7.20E-03	S		x
H368981	0.96	186	2.00E-03	*	0.05	2.30E-03	1.92E-04	***	0.07	3.44E-04	R	x	
actcaa14	0.18	183	6.00E-03	NS	0.04	1.18E-02	1.97E-04	***	0.07	6.38E-04	S		
B128918	0.32	182	5.70E-03	NS	0.04	6.40E-03	6.98E-04	**	0.06	5.70E-04	S		
B428269	0.21	177	1.50E-03	*	0.05	2.30E-03	2.70E-03	NS	0.05	5.90E-03	S		x
acactc24	0.68	93	2.00E-02	NS	0.06	1.97E-02	3.93E-04	***	0.13	4.47E-04	R		
acactc28	0.51	179	1.20E-03	*	0.05	1.80E-03	1.26E-01	NS	0.01	1.72E-01	S		
agccat11	0.82	180	4.76E-04	***	0.06	4.09E-04	3.30E-03	NS	0.04	2.20E-03	R		
agccat31	0.46	183	3.00E-03	*	0.04	4.80E-03	1.11E-01	NS	0.01	9.48E-02	S		
B385028	0.91	183	8.70E-03	NS	0.03	1.41E-02	4.57E-04	***	0.06	1.10E-03	R		
B435142	0.93	183	1.20E-03	*	0.05	1.50E-03	7.60E-03	NS	0.03	7.30E-03	R		
H252978	0.29	187	3.80E-03	*	0.04	4.10E-03	5.75E-02	NS	0.02	5.95E-02	R		
H427988	0.20	187	1.30E-03	**	0.05	1.80E-03	1.59E-02	NS	0.03	2.30E-02	S		
agcctg46	0.53	185	3.92E-02	NS	0.02	5.13E-02	1.40E-03	*	0.05	3.80E-03	R		
Number of significant associations			17				16						
Among which independent ones			14				13						
Range of R ² for significant associations			0.04-0.08				0.05-0.13						

(a) : markers mentioned in boxes are haplotype (markers linked in disequilibrium according to a significant association Fisher exact test at Bonferroni threshold)

(b) : ***, **, * significant associations with GLM model at a respective genome-wise type I error rate (GWER) of P<0.001, P<0.01 and P<0.05 tests inferred from 2000 permutations

(c) : comparative detection with MLM model (taking into account structure and kindship information) (d) : marker effect in T1 and T2 trails : R=resistance S=susceptibility

Table 3: Results of multiple linear regression of *DS* trait on independent markers significant ($P_{\text{GWER}} < 0.05$) in each trial (T1, T2)

Trial	With all markers			Best parsimonious contributive subset			
	Number of markers	df ^a	R ²	Number of markers	df ^a	R ²	List of markers
T1	19	133	0.41	9	155	0.36	actcaa14, agccat11, aggcaa33, B128918, H252978, H367877, H419516, H420097, H427561
T2	19	131	0.47	7	160	0.40	actcaa14, agccat11, B128918, B385028, H367877, H419516, H427561

(a) : Total degree of freedom of the regression model (differences between trials with all markers are due to different agronomic missing data)

References

- Aitken KS, Hermann S, Karno K, Bonnett GD, McIntyre LC, Jackson PA (2008) Genetic control of yield related stalk traits in sugarcane. *Theor Appl Genet* 117:1191-1203
- Autrey L J C, Saumtally S, Dookun A, Sullivan S, Dhayan S (1995) Aerial transmission of the leaf scald pathogen, *Xanthomonas albilineans*. *Proc Int Soc Sugar Cane Technol* 21:508-526
- Bechet GR, Bailey RA, Nuss J (1992) Screening of new sugarcane clones for resistance to leaf scald (*Xanthomonas albilineans*). *Proc S Afr Sugar Technol Assoc Annu Congr* 65:82-86
- Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinforma* 23:2633-2635
- Comstock J C (2001) Foliar symptoms of sugarcane leaf scald. *Sugar J* 64:23-32
- Champoiseau P, Daugrois JH, Girard JC, Royer M, Rott P (2006) Variation in albicidin biosynthesis genes and in pathogenicity of *Xanthomonas albilineans*, the sugarcane leaf scald pathogen. *Phytopathol* 96:33-45
- Champoiseau P, Rott P, Daugrois JH (2009) Epiphytic populations of *Xanthomonas albilineans* and subsequent sugarcane stalk infection are linked to rainfall in Guadeloupe. *Plant Dis* 94:339-345
- Daugrois J-H, Dumont V, Champoiseau P, Costet L, Boissière-Noc R, Rott P (2003). Aerial contamination of sugarcane in Guadeloupe by two strains of *Xanthomonas albilineans*. *Eur J Plant Pathol* 109:445-458
- Davis M J, Rott P, Warmuth CJ, Chatenet M, Baudin P. (1997) Intraspecific genomic variation within *Xanthomonas albilineans*, the sugarcane leaf scald pathogen. *Phytopathol* 87:316-324
- D'Hont A, Grivet L, Feldmann P, Rao S, Berding N, Glaszmann JC (1996) Characterization of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. *Mol Gen Genet* 250:405-413
- D'Hont A, Ison D, Alix K, Roux C, Glaszmann JC (1998) Determination of basic chromosome numbers in the genus *Saccharum* by physical mapping of ribosomal RNA genes. *Genome* 41:221-225
- Grivet L, Arruda P (2001) Sugarcane genomics: depicting the complex genome of an important tropical crop. *Curr Opin in Plant Biol* 5:122-127
- Gu K, Yang B, Tian D, Wu L, Wang D, Sreekala C, Yang F, Chu Z, Wang GL, White F, Yin Z (2005) R gene expression induced by a type-III effector triggers disease resistance in rice. *Nature* 435:1122-1125
- Hoarau JY, Offmann B, D'Hont A, Risterucci A-M, Roques D, Glaszmann J-C, Grivet L (2001) Genetic dissection of a modern sugarcane cultivar (*Saccharum* spp.). I. Genome mapping with AFLP markers. *Theor Appl Genet* 103:84-97
- Hoarau JY, Grivet L, Offmann B, Raboin LM, Diroflar JP, Payet J, Hellmann M, D'Hont A, Glaszmann JC (2002) Genetic dissection of a modern sugarcane cultivar (*Saccharum* spp.). II. Detection of QTLs for yield components. *Theor Appl Genet* 105:1027-1037
- Hoarau JY, Souza G, D'Hont A, Menossi M, Pinto LR, Pereira de Souza A, Grivet L, Menck CF, Ulian EC, Vincentz M (2007) Sugarcane, a tropical crop with a highly complex genome. In: *Functional Plant Genomics*, J F Morot-Gaudry, P Lea and JF Briat (Eds). Enfield NH, USA, Science Publishers, pp 481-499
- Holder DG, Irey MS, Davidson RW, Hu CJ (2000) Yield losses caused by leaf scald in a commercial field of sugarcane cultivar CP 80-1743. *J Am Soc Sugar Cane Tech*, 20:88-92
- Hoy JW (1994) Sugarcane leaf scald distribution, symptomatology and effect on yield in Louisiana. *Plant Dis* 78:1083-1087
- Heller-Uszynska K, Uszynski G, Huttner E, Evers M, Carlig J, Caig V, Aitken K, Jackson P, Piperidis G, Cox M, Arrays Gilmour R, D'Hont A, Butterfield M, Glaszmann JC, Kilian A (2011) Technology effectively reveals DNA polymorphism in a large and complex genome of sugarcane Diversity Arrays Technology (DArT) effectively reveals DNA polymorphism in a large and complex genome of sugarcane. *Mol. Breed* 8:37-55

- Jannoo N, Grivet L, Dookun A, D'Hont A, Glaszmann JC (1999) Linkage disequilibrium among modern sugarcane cultivars. *Theor Appl Genet* 99:1053-1060
- Koike H (1965) The aluminum cap method for testing sugarcane varieties against leaf scald disease. *Phytopathol* 55:317-319
- Lu Y, D'Hont A, Paulet F, Grivet L, Arnaud M, Glaszmann JC (1994) Molecular diversity and genome structure in modern sugarcane varieties. *Euphytica* 78:217-226
- Machado GR (2000) Sugarcane variety notes, an international directory 7th revision, Piracicaba, Brazil, 132 pages
- Nibouche S, Raboin LM, Hoarau JY, D'Hont A, Costet L (in press) Quantitative trait loci for sugarcane resistance to the spotted stem borer *Chilo sacchariphagus*. *Mol Breed* DOI:10.1007/s11032-010-9531-0
- Offmann B (2000) Caractérisation et analyse génétique de la résistance de la canne à sucre à *Xanthomonas albilineans*, PhD dissertation, University of Reunion island, 116 pages
- Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genet* 2:2074-2092
- Piperidis N, Jackson P, D'Hont A, Besse P, Hoarau JY, Courtois B, Aitken KS, McIntyre L (2008) Comparative genetics in sugarcane enables structured map enhancement and validation of marker-trait associations. *Mol Breed* 21:233-247
- Piperidis G, Piperidis N, D'Hont A (2010) Molecular cytogenetic investigation of chromosome composition and transmission in sugarcane. *Mol. Genet and Genomics* 284:65-73
- Pritchard JK, Stephens M, Donnelly P. (2000). Inference of population structure using multilocus data. *Genetics* 155:945-959
- Raboin LM, Pauquet J, Butterfield M, D'Hont A, Glaszmann JC (2008) Analysis of genome-wide linkage disequilibrium in the highly polyploidy sugarcane. *Theor Appl Genet* 116:701-714
- Reffay N, Jackson PA, Aitken S, Hoarau JY, D'Hont A, Besse P, McIntyre L (2005) Characterization of genome regions incorporated from an important wild relative into Australian sugarcane. *Mol Breed* 15:367-381
- Rice (1989) Analyzing tales of statistical tests. *Evolution* 43:223–225
- Ruiz M, Rouard M, Raboin LM, Lartaud M, Lagoda P, Courtois B (2004) Tropgene-DB, a multitropical crop information system. *Nucleic Acids Res* 32: D364–D367
- Rott P, Soupa D, Brunet Y, Feldmann P, Letourmy P (1995) Leafscald (*Xanthomonas albilineans*) and its effect on yield in seven sugarcane cultivars in Guadeloupe. *Plant Pathol* 44: 1075-1084
- Rott P, Davis MJ (2000) Leaf scald. In: P Rott, R A Bailey, JC Comstock, BJ Croft, AS Saumtally (eds), *A Guide to Sugarcane Diseases*. La Librairie du Cirad, Montpellier, France, pp 38-44
- SAS Institute (2008) SAS OnlineDoc® 9.2. SAS Institute Inc., Cary, NC, USA
- Walker DIT (1987) Breeding for resistance. In : Heinz D.J. (ed.), *Sugarcane Improvement through Breeding*, Elsevier Science Publishers BV, pp 445-502
- Wei X, Jackson P, McIntyre C, Aitken K, Croft B (2006) Associations between DNA markers and resistance to diseases in sugarcane and effects of population substructure. *Theor Appl Genet* 114:155-164
- Wei X, Jackson PA, Hermann S, Kilian A, Heller-Uszynska K, Deomano E (2010) Simultaneously accounting for population structure, genotype by environment interaction, and spatial variation in marker-trait associations in sugarcane. *Genome* 53:973-981
- Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genet* 38:203-208
- Zhu C, Yu J (2009) Nonmetric multidimensional scaling corrects for population structure in association mapping with different sample types. *Genetics* 182:875-888

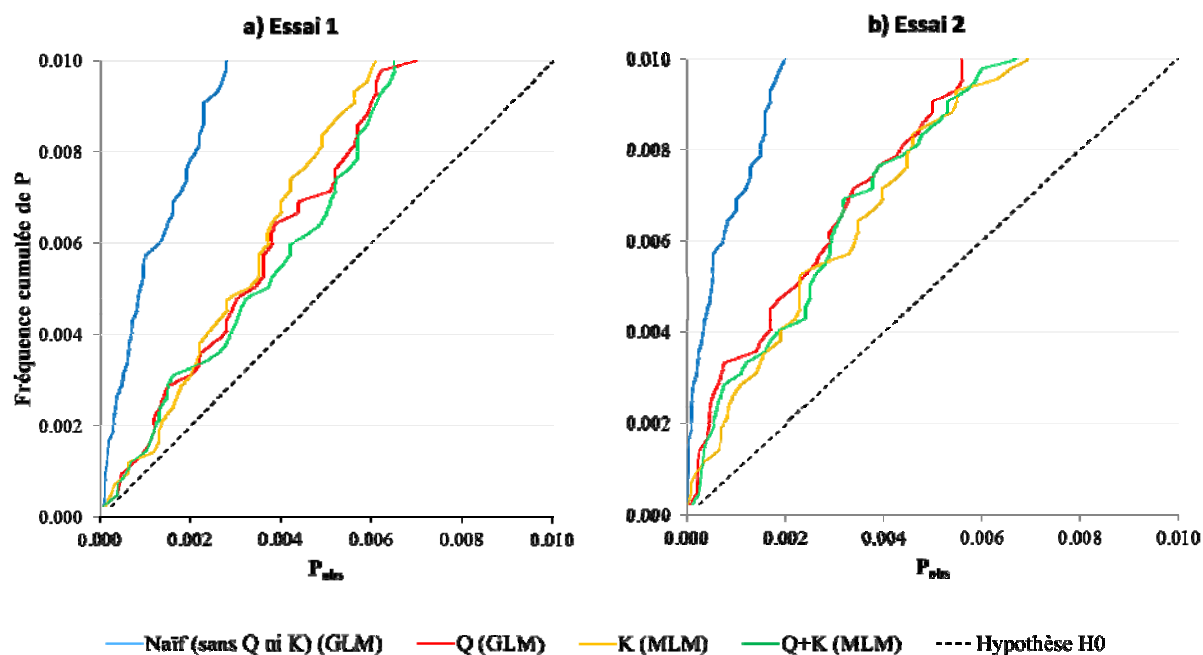


Figure 18 : Distribution cumulée des p-value observées (P_{obs}) par rapport à une loi de probabilité uniforme, des différents modèles d'associations utilisés (Naïf, Q, K, Q+K), concernant la résistance à la propagation aérienne de l'échaudure des feuilles. Sous l'hypothèse d'une absence de marqueurs liés au polymorphisme contrôlant la résistance à la propagation aérienne de l'échaudure des feuilles (H_0), le modèle d'analyse qui contrôle le mieux les erreurs de type I tend vers une distribution uniforme des P-values.

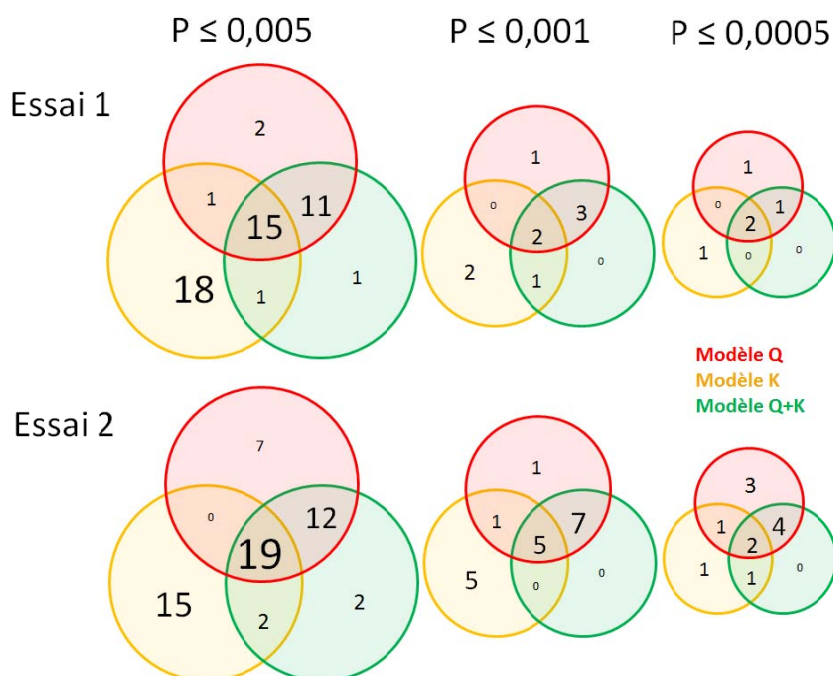


Figure 19 : Nombre d'associations détectées par chacun des modèles (Q, K et Q+K) pour la résistance à la contamination aérienne à l'échaudure des feuilles, dans les deux essais, à différentes P-values nominales.

I.3 Données complémentaires

I.3.1 Comparaison des différents modèles d'associations

La distribution cumulée des p-values nominales observées pour l'ensemble des 4189 marqueurs a été étudiée pour chaque essai, afin de comparer les résultats d'associations obtenus avec les différents modèles utilisés (Naïf, Q, K et Q+K). La **Figure 18** montre le comportement des modèles pour des p-values inférieures à 0,01.

Les 4 modèles montrent des courbes similaires entre les deux essais, toutes situées au-dessus de la bissectrice, indiquant qu'il existe plus d'associations marqueur-caractère que le nombre attendu par le simple fait du hasard statistique. Le modèle naïf, qui ne prend en compte ni structure ni apparentement entre individus, implique la détection d'un grand nombre d'associations dont une forte proportion est potentiellement artéfactuelle (111 et 121 associations décelées respectivement dans les essais 1 et 2 au seuil des p-values nominales inférieur à 0,01 défini par la figure). Les trois courbes correspondant aux modèles Q, K, et Q+K, qui se superposent dans les deux essais, indiquent un nombre d'associations comparable entre les 3 modèles, qui est deux fois inférieur à celui du modèle naïf au seuil nominal de 0,01. Plusieurs études d'associations ont montré que les modèles d'analyses fondés sur la prise en compte de co-facteurs représentatifs de la structure et d'apparentement sont efficaces pour limiter le risque d'inflation d'associations lié à un exercice de type « genome scan » (Yu *et al.*, 2006 ; Zhao *et al.*, 2007 et 2011 ; Aranzana *et al.*, 2005). La division par deux du nombre d'associations que nous constatons dans cette étude d'association relative à la résistance à l'échaudure des feuilles confirme la validité de l'emploi de ces méthodologies d'analyses dans le contexte génétique particulier de la canne à sucre, visant à limiter les associations faussement positives (Wei *et al.*, 2010).

La superposition des courbes des 3 modèles testés (Q, K, Q+K), qui traduit un nombre d'associations comparables à p-values nominales identiques, cache en fait des répertoires d'associations qui se recouvrent plus ou moins bien selon les modèles. Ces recouvrements partiels s'expliquent par l'influence différente exercée par les co-facteurs pris en compte dans les trois modèles d'analyse (Q, K, Q+K) et sont en accord avec les résultats observés par plusieurs auteurs (Achleitner *et al.*, 2008 ; Stracke *et al.*, 2009).

Comme le montre la **figure 19**, qui illustre le nombre d'associations communes entre modèles observables sur une gamme de p-values nominales, la tendance est la même sur les 2 essais dans toutes les situations de p-values : les modèles Q et Q+K partagent toujours plus de marqueurs en commun que n'en partagent les modèles Q et K ou bien les modèles K et Q+K. Ces résultats suggèrent que le facteur représentatif de la structure (Q) apporte une contribution plus importante que le co-facteur de kinship (K) dans le résultat du modèle global (K+Q) qui est théoriquement le plus performant. Cela étant dit, les associations que ces modèles détectent ne le sont que sur la base de probabilités nominales. Comme développé dans l'article, seules les associations ultérieurement validées à travers un deuxième filtre statistique (GWER) qui contrôle le risque de 1^{ère} espèce sur l'ensemble du génome permet d'identifier *in fine* les marqueurs les plus robustes contribuant à la variation des phénotypes observés.

Comme indiqué dans l'article, la ségrégation agronomique du premier essai (moy = 0,19 ; ET = 0,17) est légèrement plus large que celle de l'essai 2 (moy = 0,11 ; ET = 0,13). De surcroît, pour les modèles Q et Q+K, le co-facteur représentant la structure Q est fondé sur la prise en compte d'un nombre d'axes ACP significatifs plus élevé dans le deuxième essai (7 axes expliquant 15 % de la variation phénotypique) que dans le premier (4 axes expliquant 11 % de la variation phénotypique). La variation agronomique restante une fois soustraite la part de variation attribuée à la structure (Q) est donc d'autant plus réduite dans l'essai 2. Sachant que la précision du phénotypage est comparable dans les deux essais ($H^2_{E1}=0,89$; $H^2_{E2}=0,83$), la ségrégation plus large de l'essai 1 a pu permettre une limitation un peu meilleure du nombre de faux positifs révélé à des probabilités nominales. L'écart à la bissectrice des trois courbes (Q, K, Q+K) qui est légèrement plus prononcé dans l'essai 2 que dans l'essai 1 semble confirmer cette hypothèse.

I.3.2 Détermination de l'origine ancestrale spécifique des marqueurs

La détermination de l'origine spécifique (*S. officinarum* versus *S.spontaneum*) des marqueurs produits dans notre étude d'association nécessite de comparer le profil de deux panels d'accessions représentatifs de chacune des 2 espèces ancestrales. Pour réaliser cette étude, l'ADN de 47 clones « *S. officinarum* » et de 32 clones « *S. spontaneum* » était

Tableau 7 : Liste et provenance des clones ancestraux *S. officinarum* et *S. spontaneum* utilisés pour la détermination de l'origine ancestrale spécifique des marqueurs DArT.

Clones déclarés d'origine <i>S. officinarum</i>			Clones déclarés d'origine <i>S. spontaneum</i>		
Code	Origine de l'échantillon	Nom clone	Code	Origine de l'échantillon	Nom clone
o1	Collection MPL	Ashy Mauritius	s1	Afrique du Sud	Coimbatore
o2	Maurice	BA11569	s2	Collection MPL	Coimbatore local
o3	Collection MPL	Bambou rose	s3	Collection MPL	Ghagha kep
o4	Coll. Réunion	Batjan	s4	Coll. Réunion	Glagha Wt
o5	Afrique du Sud	Black cheribon	s5	Coll. Réunion	Glagha1286
o6	Collection MPL	Crystalina	s6	Collection MPL	IK76-67
o7	Collection MPL	D 109	s7	Coll. Réunion	Mandalay
o8	Coll. Réunion	D74	s8	Afrique du Sud	Pasoeroean
o9	Collection MPL	EK2	s9	Guadeloupe	SES14
o10	Collection MPL	Fidji24	s10	USA	SES208A
o11	Coll. Réunion	Lousier	s11	Collection MPL	IK76-86
o12	Collection MPL	Mauritius Ghunghan	s12	Coll. Réunion	Coimbatore aus
o13	Coll. Réunion	Oramboo	s13	Collection MPL	IK7641
o14	Collection MPL	Vellai	s14	Collection MPL	Coimbatore
o15	Collection MPL	BNS3066	s15	Collection MPL	Glagha Wt
o16	Collection MPL	Fiji44	s16	Collection MPL	Glagha1286
o17	Collection MPL	IJ76291	s17	Collection MPL	Isolo
o18	Collection MPL	IK7633	s18	Collection MPL	Kloet
o19	Barbades	BA11569	s19	Collection MPL	Mandalay
o20	Afrique du Sud	Badila	s20	Collection MPL	Pasoeroean
o21	USA	Banjarmasin Hitam	s21	Coll. Réunion	SES14
o22	USA	Black cheribon	s22	Collection MPL	SES178
o23	Coll. Réunion	Crystalina	s23	Collection MPL	SES208A
o24	Afrique du Sud	D74	s24	Collection MPL	SES264
o25	Collection MPL	Korpi	s25	Réunion	Tabongo
o26	USA	LA Purple	s26	Afrique du Sud	US 1694
o27	Brésil	Lahaina	s27	Collection MPL	US56158
o28	Collection MPL	Loethers	s28	Collection MPL	Mol5801
o29	Collection MPL	Lousier	s29	Collection MPL	Mol5904
o30	Coll. Réunion	Mauritius Ghunghan	s30	Collection MPL	NG51-2
o31	Collection MPL	NG5752	s31	Collection MPL	Tainan96
o32	Brésil	Oramboo	s32	Collection MPL	US560093
o33	Brésil	Otaheite			
o34	Collection MPL	POJ100			
o35	USA	Stripped preanger			
o36	Brésil	Vellai			
o37	Collection MPL	NC29			
o38	Collection MPL	NC42			
o39	Collection MPL	NC99			
o40	Collection MPL	NC24			
o41	Collection MPL	NC30			
o42	Collection MPL	NG57198			
o43	Collection MPL	IS76116			
o44	Collection MPL	Baltazar			
o45	Réunion	Formosa			
o46	Collection MPL	chittan			
o47	Collection MPL	Kaludai Boothan			

Clones d'origine potentiellement interspécifique au vue de l'analyse de diversité

Clone d'origine ancestrale potentiellement éronnée

disponible (**Tableau 7**). Ces ADN provenant de collectes variées (Afrique du Sud, Brésil, Barbade, Ile Maurice, USA, la Réunion, Guadeloupe, Montpellier), nous avons d’abord entrepris une vérification de la légitimité de leur origine ancestrale déclarée sur la base d’une étude de diversité génétique de l’ensemble de ces 79 accessions. Presque tous les principaux clones *S. officinarum* connus pour avoir contribué à la généalogie des cultivars modernes (Arceneaux, 1965) figurent dans le pool des 47 accessions : Black Chérifon (o5, o22), Bandjarmasin Hitam (o21), Loethers (o28), Crystalina (o6, o23), Ashy Mauritius (o1), Kaludai Boothan (o47), Vellai (o14, o36) ; Badila (o20), Lahaina (o27), Striped Preanger (o35), Korpi (o25).

a. Etude de diversité pour l’identification sécurisée de deux panels de clones ancestraux

Les 79 clones ancestraux (**Tableau 7**) ont été génotypés sur les DArT en même temps que les populations d’association de Guadeloupe (189 cultivars) et de la Réunion (184 cultivars). L’étude de leur diversité présentée ici a été réalisée avec 625 marqueurs DArT représentant un sous-ensemble des 867 DARTs indépendants identifiés dans la population d’association de Guadeloupe. L’arbre d’apparentement entre les 79 accessions présenté à la **Figure 20(a)** a été établi avec le logiciel DARwin (Perrier and Jacquemoud-Collet, 2006) sur la distance euclidienne moyenne entre chaque couple (i,j) d’individus

$$[d_{ij} = \sqrt{\frac{1}{k} \sum_1^K (x_{ik} - x_{jk})^2}].$$

La **Figure 20 (a)** révèle que le clone s13, considéré comme appartenant à l’espèce *S. spontaneum*, est sans doute un clone *S. officinarum*. Ce clone a donc été exclu de la constitution sécurisée des deux panels ancestraux recherchés. Par ailleurs, 14 autres clones (2 « *S. officinarum* » et 12 « *S. spontaneum* »), montrent une position intermédiaire entre les deux pôles *S. officinarum* et *S. spontaneum* et pourraient, pour certains d’entre eux, être des hybrides interspécifiques (o3, o4, o8, o10, o19, o21, o22, o23, o24, o34, o45, o46, o26, o28). Par sécurité ces 14 clones ont donc aussi été délaissés. Après ces éliminations il subsiste un groupe de 35 clones *S. officinarum* et de 29 clones *S. spontaneum* suffisamment distants

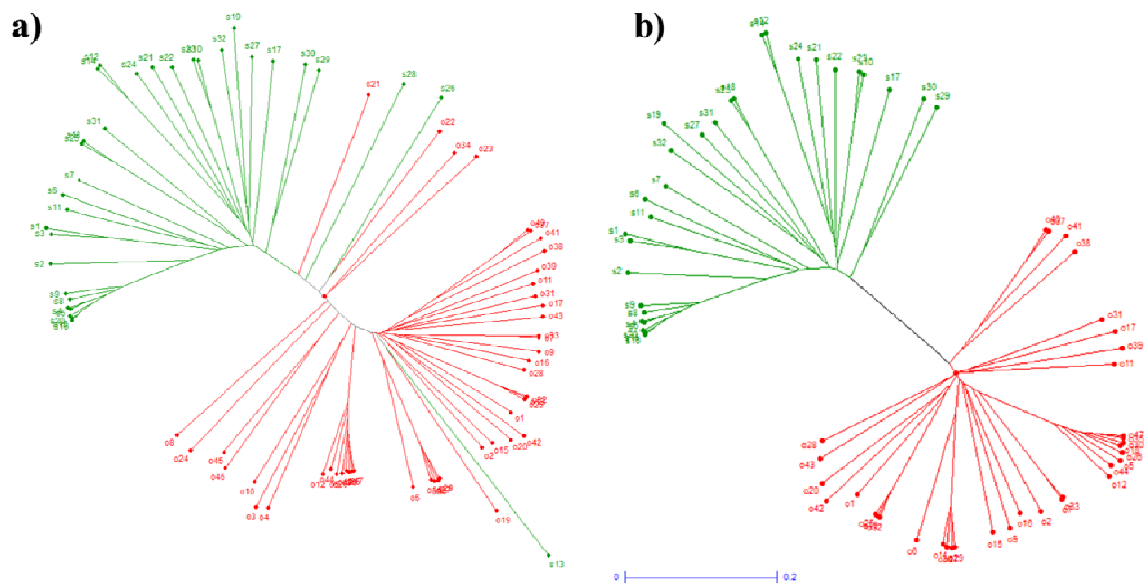


Figure 20 : Arbres de diversité représentant la distance euclidienne moyenne entre clones ancestraux, calculée à l'aide de 625 marqueurs DArT codés dans la population d'association de Guadeloupe. Les clones d'origine ancestrale supposée *S. officinarum* sont représentés en rouge, ceux d'origine ancestrale supposée *S. spontaneum* en vert. a) Ensemble des 79 clones ancestraux génotypés b) Sous-ensembles de 64 clones ancestraux considérés comme légitimes et utilisés pour l'assignation des marqueurs DArT à une espèce ancestrale.

génétiqnement pour entreprendre une détermination de l'origine spécifique des marqueurs DArT limitant les risques d'assignations erronées (**Figure 20 (b)**).

b. Origine spécifique des marqueurs DArT

Compte tenu du nombre limité de clones *S. officinarum* impliqués dans la généalogie des cultivars modernes, et du taux de polymorphisme modeste existant dans cette espèce ancestrale associé à un niveau d'hétérozygotie élevé (Lu *et al.*, 1994), l'analyse d'un nombre limité de clones *S. officinarum* devrait permettre d'identifier la majorité des allèles issus de cette espèce existant dans les cultivars interspécifiques modernes.

L'origine ancestrale des marqueurs DArT a été déterminée en considérant les règles suivantes :

- 1) un marqueur est d'origine *S. officinarum* s'il est présent chez au moins un des 35 clones *S. officinarum*, et absent chez tous les 29 individus *S. spontaneum* ;
- 2) inversement, un marqueur est d'origine *S. spontaneum* s'il est présent chez au moins un des 29 individus *S. spontaneum*, et absent chez tous les 35 individus *S. officinarum* ;
- 3) tout marqueur polymorphe dans chacun des 2 pools *S. officinarum* et *S. spontaneum* (présent chez au moins un individu) n'est spécifique d'aucune des deux espèces.

Sur les 1713 marqueurs DArT présents dans la population d'association, 1282 ont été codés chez les 64 clones ancestraux retenus. Parmi ces 1282 marqueurs 26 (2 %) sont absents chez tous les 64 clones, 132 (10 %) sont d'origine *S. officinarum*, 527 (41 %) sont d'origine *S. spontaneum*, et 597 (47 %) ne sont spécifiques d'aucune des deux origines (présents chez les deux espèces à la fois).

La faible proportion de marqueurs « monomorphes absents » (2%) indique que les deux panels ancestraux utilisés couvrent presque toute la diversité allélique présente dans la population d'associations de Guadeloupe (189 cultivars).

Dans le génome des cultivars modernes, 73 à 85 % des chromosomes sont entièrement d'origine *S. officinarum*, 10 à 23 % d'entre eux sont uniquement d'origine *S. spontaneum* et 8 à 13 % des chromosomes sont des recombinants entre les deux espèces ancestrales (D'Hont *et al.*, 1996 ; Piperidis *et al.*, 2010). Nous attribuons ici quatre fois plus

Tableau 8 : Origine ancestrale des marqueurs associés à la propagation aérienne de l'échaudure des feuilles ($P_{GWER} < 0,05$). (Type D : marqueur DArT, A : marqueur AFLP ; Effet S: marqueur associé à la sensibilité, Type R: marqueur associé à la résistance).

Type	Maqueur	Effet	Origine ancestral
D	H420097	S	<i>S. spontaneum</i>
D	B420097	S	<i>S. spontaneum</i>
A	aggcaa33	S	<i>S. spontaneum</i>
A	aagctt28	S	-
D	H382397	R	<i>S. spontaneum</i>
A	acccat2	R	-
D	H257517	S	Non spécifique
D	H370204	S	Non spécifique
D	H367877	S	Non spécifique
D	H416781	S	Non spécifique
D	H370244	S	Non spécifique
D	H427561	S	<i>S. spontaneum</i>
A	acccac54	S	<i>S. spontaneum</i>
A	aggcac11	R	<i>S. spontaneum</i>
A	aggctt1	R	-
D	H419516	S	-
D	H368981	R	-
A	actcaa14	S	-
D	B128918	S	-
D	B428269	S	-
A	acactc24	R	-
A	acactc28	S	<i>S. spontaneum</i>
A	agccat11	R	-
A	agccat31	S	-
D	B385028	R	Non spécifique
D	B435142	R	Non spécifique
D	H252978	R	Non spécifique
D	H427988	S	-
A	agcctg46	R	-

de marqueurs DArT de la population d'association de Guadeloupe (189 cultivars) à une origine *S. spontaneum* (41 %) qu'à une origine *S. officinarum* (10 %). Ces proportions de marqueurs attribuées aux deux origines spécifiques sont donc inversement corrélées au poids de leur génome dans les cultivars modernes. Ces résultats DArT montrent qu'il existe parmi les cultivars étudiés un niveau de polymorphisme supérieur dans le sous-génome hérité de *S. spontaneum* comparativement au sous-génome provenant de *S. officinarum*. Ils sont conformes à la carte génétique de référence (cv. R570) qui montre une meilleure couverture des chromosomes hérités de l'espèce sauvage (Grivet *et al.*, 1996 ; Hoarau *et al.*, 2001 ; Rossi *et al.*, 2003 ; données non publiées). Cela s'explique par le nombre très élevé de chromosomes *S. officinarum* (8 à 9) dans chaque classe d'homéologie (Grivet et Arruda, 2001) qui se traduit par le fait que l'on accède plus difficilement au polymorphisme contenu dans ce sous-génome ancestral. Nos résultats DArT confirment donc la contribution importante de l'espèce sauvage à la diversité génétique des cultivars modernes (Lu *et al.*, 1994) ou tout au moins à la fraction de cette diversité accessible avec les techniques de marquage actuelles.

Parmi les 23 associations indépendantes détectées ($P_{\text{GWER}} < 0,05$) pour la résistance à l'échaudure des feuilles sur l'ensemble des deux essais agronomiques (Table 2 de l'article, p56), 13 associations impliquent des marqueurs DArT (**Tableau 8**). Parmi ces 13 détections, 3 sont d'origine *S. spontaneum*, 5 ne sont spécifiques d'aucune des 2 espèces ancestrales, et l'origine des 5 autres marqueurs n'est pas déterminée en l'absence de codage disponible.

c. Origine spécifique des marqueurs AFLP

Pour des contraintes techniques (disponibilité limitée de pistes de dépôt), seules 14 accessions « *S. officinarum* » ont été génotypées en AFLP en même temps que la population d'association de Guadeloupe : Ashy Mauritius (o1), Batjan (o4), BA11569 (o19), Banjarmasin Hitam (o21), Badila (o20), D109 (o7), La Purple (o26), Loethers (o28), POJ 100 (o34), Kaludai Boothan (o47), Black Chéribon (o5 ou o22), Crystalina (o6 ou o23), DK74 (o8 ou o24), Vellai (o14 ou o36).

Ce choix de ne génotyper qu'un petit nombre d'accessions *S. officinarum* a été motivé par le fait que la diversité relativement faible au sein de cette espèce (Lu *et al.*, 1994) offre de bonnes chances « d'assigner par défaut » une origine *S. spontaneum* à tout marqueur qui serait totalement absent du petit groupe de clones *S. officinarum* (Hoarau *et al.*, 2001) ; le raisonnement inverse avec un petit groupe de clones *S. spontaneum* n'étant pas concevable

compte tenu de la diversité génétique sensiblement plus large attendue au sein de cette espèce ancestrale.

Parmi les 14 accessions « *S. officinarum* » génotypées avec les marqueurs AFLP :

- 7 accessions s'avèrent être des clones *S. officinarum* relativement certains [Ashy Mauritius (o1), Badila (o20), D109 (o7), La Purple (o26), Loethers (o28), Vellai (o14 ou o36), Kaludai Boothan (o47)] ;
- 2 accessions [Black Chérignon (o5 ou o22), Crystalina (o6 ou o23)] peuvent chacune correspondre à deux échantillons différents dont un s'avère être un clone *S. officinarum* certain et l'autre pourrait être d'une origine interspécifique ;
- 5 accessions font partie du groupe de clones que nous considérons a posteriori comme éventuellement d'origine interspécifique, [Batjan (o4), BA11569 (o19), Banjarmasin Hitam (o21), POJ 100 (o34), DK74 (o8 ou o24)].

Par prudence les 5 dernières accessions ont été éliminées du raisonnement d'assignation. Seules n'ont été considérées qu'un premier groupe constitué des 7 premières accessions, puis un second groupe de 9 accessions correspondant au premier groupe élargi aux deux accessions suivantes. Tout marqueur absent du premier groupe a été assigné par défaut à une origine *S. spontaneum*. Le même raisonnement a été mené avec le second groupe de clones, et les résultats d'assignation obtenus ont été comparés.

Sur les 2476 marqueurs AFLP codés sur la population d'association, 1241 marqueurs ne possédaient aucune donnée de codage manquante sur l'ensemble des 9 accessions *S. officinarum* examinées. Parmi les 1241 marqueurs ayant un codage complet 489 marqueurs (39 %) s'avèrent attribuables à une origine spécifique *S. spontaneum* quand on considère la population de 7 accessions *S. officinarum*, alors que 511 marqueurs (41 %) possèderaient une origine *S. spontaneum* si l'on se base sur la seconde population de 9 accessions *S. officinarum*. De plus, les deux raisonnements aboutissent à des assignations d'origine spécifique parfaitement concordantes des marqueurs significativement associées à la résistance à l'échaudure des feuilles (P_{GWER} de 0,05). Ces deux résultats suggèrent que les deux accessions supplémentaires (*S. officinarum*) sont légitimes. Comme le montre le tableau 8, quatre marqueurs AFLP (ou groupes de marqueurs) possèdent une origine spécifique imputable à l'espèce sauvage. Les 7 autres marqueurs AFLP ne sont pas attribuables à une origine spécifiquement sauvage. En l'absence de comparaison avec un échantillon de clones

sauvages, il n'est pas possible de déterminer, par un raisonnement d'assignation symétrique, si ces 7 marqueurs restant sont d'origine *S. officinarum*, ou bien s'il s'agit de marqueurs spécifiques d'aucune des deux origines ancestrales.

Bien que les clones *S. officinarum* soient connus pour être peu polymorphes entre eux (Lu *et al.* 1994), la relative modestie du panel de clones *S. officinarum* utilisés (9 accessions) pourrait ne pas totalement garantir l'exactitude de l'identification des marqueurs AFLP correspondant à une origine sauvage. A la différence de l'exercice précédent mené sur les DArT, les conclusions obtenues ici doivent être considérées avec prudence. Toutefois, la nature proche du polymorphisme généré par les deux techniques de marquage et la quasi-égalité des pourcentages de marqueurs spécifiquement attribuables à une origine *S. spontaneum* entre AFLP (39% et 41%) et DArT (41%) donnent du crédit à la procédure employée et aux conclusions obtenues sur les AFLP.

d. Origine des marqueurs associés au phénotype agronomique

Sur les 23 associations indépendantes détectées ($P_{\text{GWER}} < 0,05$) pour la résistance à l'échaudure des feuilles, on dénombre une majorité de 13 associations impliquant des marqueurs DArT et 10 associations fondées sur des marqueurs AFLP (**Tableau 8**).

Concernant les DArT dont l'origine spécifique a pu être examinée avec un raisonnement complet au moyen de deux panels ancestraux étoffés, on ne dénombre aucune association correspondant spécifiquement à une origine *S. officinarum*. Les 8 associations dont l'origine a pu être déterminée correspondent soit à des marqueurs *S. spontaneum* (3 associations) soit à des marqueurs non spécifiques (5 associations).

Tous types de marqueurs confondus (AFLP et DArT), on dénombre un total de 7 associations indépendantes qui relèvent d'une origine spécifiquement sauvage. Sur ces 7 allèles, 5 sont liés à un accroissement de la sensibilité, et 2 allèles à une amélioration de la résistance. Parmi les 5 associations qui ne sont spécifiques ni de *S. officinarum*, ni de *S. spontaneum*, 3 correspondent à un marqueur de résistance et 2 à un marqueur de sensibilité.

Dans l'ensemble, même s'ils sont incomplets, les résultats que nous obtenons montrent que la fraction de la variation phénotypique de la population qui a été marquée est principalement déterminée par des allèles non spécifiques, ou bien originaires de l'espèce sauvage.

Sur les marqueurs DArT où le raisonnement a été complet, il apparaît qu'aucun marqueur ayant pu être déterminé ne provient spécifiquement de l'espèce domestiquée. L'hypothèse de la présence de quelques marqueurs *S. officinarum* parmi ceux dont l'origine reste encore indéfinie, n'est pas exclue. Toutefois, les statistiques de fréquences précédemment établies (*cf.* paragraphe I.3.2.1, page 62) rendent cette hypothèse peu probable.

La diversité agronomique observée pour la contamination aérienne de l'échaudure des feuilles des cultivars modernes semble donc principalement déterminée par des allèles de sensibilité ou de résistance apportés par l'espèce sauvage (*S. spontaneum*). Cette perception est cohérente avec le fait que d'une part, le génome des cultivars modernes est majoritairement issu d'un fond génétique domestiqué et que d'autre part, la plupart des ressources *S. officinarum* sont sensibles à très sensibles à l'échaudure des feuilles et montrent peu de variabilité (Daugrois *et al.*, 1995; Egan, 1972).

**Troisième partie : Etude d'association relative à
la résistance de la canne à sucre à la feuille jaune,
maladie causée par le *Sugarcane Yellow Leaf Virus***

I.1 Introduction

La feuille jaune, causée par le SCYLV, est une maladie considérée comme émergente et l'étude de sa résistance chez l'hôte dépend d'interactions potentiellement complexes dans un pathosystème où interviennent la plante hôte, le virus et le puceron vecteur (*Melanaphis sacchari*). Le travail présenté dans ce chapitre vise à tester la faisabilité d'études d'associations relatives à la résistance à cette maladie, et à comparer les résultats obtenus sur deux larges panels de cultivars, dont l'un a été expérimenté en Guadeloupe (189 clones) et l'autre à la Réunion (184 cultivars). Chaque panel a été génotypé par plus de 4000 marqueurs polymorphes. Un ensemble de 1200 marqueurs DArT codés en commun est utilisé pour la comparaison des deux exercices d'analyses génétiques.

I.2 Publication associée

Fartek B., **Rocher S.**, Debibakas S., Hervouet C., Royaert S., Raboin L.M., Roques D., Thong Chane A., Daugrois J.H., D'Hont A., Reynaud B., Hoarau J.Y., Nibouche S. and Costet L.

Association mapping of sugarcane resistance to Sugarcane Yellow Leaf Disease.

En préparation

B. Fartek^{1,4}, S. Rocher^{2,3,8}, S. Debibakas^{6,7}, C. Hervouet², S. Royaert^{1*}, L.M. Raboin⁹, D. Roques³, A. Thong Chane⁵, J.-H. Daugrois⁶, A. D'Hont², B. Reynaud¹, J.-Y. Hoarau³, S. Nibouche¹ and L. Costet^{1*}

Association mapping of sugarcane resistance to Sugarcane Yellow Leaf Disease

¹Cirad, UMR PVBMT, F-97410 Saint-Pierre, La Réunion, France

²Cirad, UMR AGAP, F-34398 Montpellier, France

³Cirad, UMR AGAP, F-97170 Petit Bourg, Guadeloupe, France

⁴Université de la Réunion, UMR PVBMT, F-97410 Saint-Pierre, La Réunion, France

⁵eRcane, F-97494 Sainte-Clotilde, La Réunion, France

⁶Cirad, UMR BGPI, F-97170 Petit Bourg, Guadeloupe, France

⁷Université des Antilles et de la Guyane, F-97157 Pointe-à-Pitre, Guadeloupe, France

⁸Université Montpellier 2, F-34095, Montpellier, France

⁹Cirad, UPR SCA, MDG-110 Antsirabe, Madagascar

*Corresponding author: Laurent Costet, Cirad, UMR PVBMT, Pôle de Protection des Plantes, 7 chemin de l'IRAT, 97410 Saint-Pierre, île de la Réunion, France. E-mail: laurent.costet@cirad.fr

Abstract

Sugarcane Yellow Leaf is a major viral disease of sugarcane caused by a polerovirus - the *Sugarcane yellow leaf virus* (SCYLV) – and transmitted by several aphid species. Varietal resistance against the virus or its vectors is the most practicable control method against Yellow Leaf disease. We analyzed the genetic basis of sugarcane resistance to SCYLV and to its aphid vector *Melanaphis sacchari*, based on association mapping using 338 cultivars. These cultivars were separated in two panels, each panel was tested either in Reunion or in Guadeloupe, presenting two different epidemiological contexts. Both panels were evaluated for SCYLV incidence in field trials under natural infection conditions. The Reunion panel was also evaluated for *M. sacchari* infestation. Both panels were genotyped with AFLP, SSR and DArT markers. Genetic structure of each panel was analyzed by Principal Component

Analysis. Analysis of marker-trait associations for resistance to SCYLV, using a mixed linear model, revealed 21 markers grouped in 12 independent haplotypes in Reunion, and 32 markers grouped in 19 independent haplotypes in Guadeloupe. Among the haplotypes common to both panels, five were detected in both panels/environments. Analysis of marker-trait associations for resistance to aphids detected 25 significant markers grouped in 18 independent haplotypes. Three of these haplotypes were significantly associated with reduction of SCYLV incidence in Reunion or in Guadeloupe.

Key words: Association mapping, *Polerovirus*, *Melanaphis sacchari*, resistance, *Saccharum* spp.

Introduction

Yellow leaf disease of sugarcane has a major economic importance in Reunion, in Guadeloupe and in most sugarcane producing areas in the world (Rott *et al.* 2008). This disease has been described recently (Schenck 1990, 1997) and causes important yield losses in susceptible cultivars, even when infected plants do not exhibit disease symptoms. It is caused by the *Sugarcane yellow leaf virus* (SCYLV), a persistent circulative nonpropagative polerovirus vectored by aphids (D'Arcy and Domier 2005; Vega *et al.* 1997). In Reunion and Guadeloupe the virus is mainly transmitted by the aphid *Melanaphis sacchari* (Zehntner) (Rassaby *et al.* 2004, Daugrois *et al.* 2011). Six genotypes of the SCYLV have been described worldwide: BRA, CUB, PER, REU (Abu Ahmad *et al.* 2007; Abu Ahmad *et al.* 2006a), IND (Viswanathan *et al.* 2008) and CHN1 (Wang and Zhou 2010). BRA-PER (the available primer set cannot distinguish between the two genotypes), REU and CUB genotypes are present in Guadeloupe, whereas only BRA-PER and REU are present in Reunion (Abu Ahmad *et al.* 2007; Abu Ahmad *et al.* 2006a). In Reunion, the sugarcane yellow leaf disease is responsible for heavy yield losses, reaching 37% in first ratoon for the cultivar R577 (Rassaby *et al.* 2003). Improvement of varietal resistance would be of great economic interest. However, because screening for resistance based on the symptoms is difficult and because the genetic basis of resistance is unknown, breeding progress to increase resistance is difficult. Indexation of sugarcane and related species germplasm collections have showed that modern cultivars (*Saccharum* spp.), *S. officinarum*, *S. robustum*, *S. sinense* present high virus incidence, whereas for *S. barberi*, *S. spontaneum*, *Miscanthus* spp. and *Erianthus* spp. present

low virus incidence (Comstock *et al.* 2001; Komor 2010; Schenck and Lehrer 2000). These last groups represent potential sources of resistance to yellow leaf disease. Resistance level to *M. sacchari* was recently analyzed for five commercial cultivars (Akbar *et al.* 2010) and exhibited genotypic variability.

The potential of linkage disequilibrium (LD)-based association studies to identify marker-trait associations has been highlighted early in sugarcane (Jannoo *et al.* 1999). The extent of LD observed among modern sugarcane cultivars is the consequence of the recent breeding history of this crop. All modern cultivars are derived from a few cross performed a century ago between the sugar producing *S. officinarum* and the wild *S. spontaneum*. The low number of generations (<10) that separates modern cultivars from the small core of sugarcane founders they derived from, has limited the number of meiosis and consequently chromosome recombination events. LD generally drops when markers are 5 cM or more apart but some instances LD can be found between markers separated by up to 40 cM (Raboin *et al.* 2008; Raboin 2005). However, many instances of linked markers cannot be recognized, mainly due to the confounding effect of polyploidy. Several studies have applied association mapping in sugarcane allowing to detect association between markers and several traits including resistance to smut, African stalk borer, pachymetra root rot, leaf scald, Fiji leaf gall (Raboin 2005, Wei *et al.* 2006, Butterfield 2007) and yield components (Wei *et al.* 2010).

The objective of this study was to use association mapping approach on two panels of modern sugarcane cultivars to identify markers associated with incidence of SCYLTV and incidence of the aphid vector *Melanaphis sacchari*.

Material and methods

Plant material

Two panels representing a total of 344 sugarcane cultivars originating from more than 30 breeding stations around the world were selected for this study. The first panel, evaluated in Reunion, consisted of 184 cultivars. The second one, evaluated in Guadeloupe, consisted of 189 cultivars, from which 29 were common with the Reunion panel.

Field trials

In Reunion, cultivars were compared in an alpha-lattice design planted at Cirad Saint-Pierre Station in 2006 with three replications, 20 blocks per replication and 10 cultivars per block. Elementary plots consisted of one 2.5 m long row and were 1.5 m wide. To reduce interaction between adjacent plots and to maximise aphid infestation and virus infection, each elementary plot was separated from the neighbouring plot by a row planted with the cultivar B 34/104 which is susceptible to aphids and SCYLV.

In Guadeloupe (at Cirad Roujol station), cultivars were compared in two trials (T1 and T2) using randomized complete block designs with three replications, planted respectively in 2005 and 2007. Elementary plot size was similar as in Reunion.

SCYLV incidence assessment

SCYLV detection was performed on the first fully emerged leaf by tissue-blot immunoassay (TBIA) as described by Schenck *et al.* (1997). A stereo-microscope (x100) was used to examine each midrib print on the TBIA membranes. A stalk was considered infected by SCYLV when at least one vascular bundle showed a positive reaction. For the Reunion panel, ten leaves from ten different stalks randomly distributed in each elementary plot were sampled in the second ratoon crop-cycle in 2009. For the Guadeloupe panel, ten to 15 leaves from different stalks randomly distributed in each plot were sampled in the second ratoon crop-cycle. Sampling was realised in 2007 for T1 and 2010 for T2.

An estimation of SCYLV prevalence regarding each of the three viral genotypes (BRA-PER, CUB and REU) present in infected cultivars in Guadeloupe and Reunion was performed in each panel by using RT-PCR methods with specific primers described by Abu Ahmad *et al.* (2006b). In Reunion, for each cultivar scored as infected by TBIA, one of the infected leaves was randomly chosen and tested by RT-PCR. In Guadeloupe, viral genotype identification was carried out on a set of 50 cultivars randomly sampled in trial T2: 30 leaves per cultivar (ten per elementary plot in each replication) were bulked prior to RNA extraction and RT-PCR analysis.

Aphid incidence recording

Aphid infestation was evaluated in the Reunion trial. Aphid incidence was scored every two weeks during 14 weeks for the 2007-2008 cropping season, 20 weeks for the 2008-2009 cropping season, and 24 weeks for the 2009-2010 cropping season, for a total of 29 records. At each observation date, in each elementary plot, the lowest green leaf of 20 randomly selected stalks in each elementary plot was observed. A leaf was recorded as infested when at least one aphid was observed on it, and the percentage of infested leaves per plot was computed.

Genotyping

AFLP analysis was performed using the AFLP® Analysis System I (Invitrogen) as recommended by the manufacturer with slight modifications. Thirty-six and 59 primer pair combinations were used for the Reunion and Guadeloupe panels, respectively. AFLP analysis was performed as described in Hoarau *et al.* (2001). Radioactive labeling was used for the Guadeloupe panel and the AFLP fingerprints were scored visually on autoradiogram. Fluorescent labeling was used for the Reunion panel and electrophoresis was performed in a 3130xl Genetic Analyzer (Applied Biosystems). The AFLP fingerprints were analyzed visually using GelCompar II software (Applied Maths BVBA). For SSR analysis, two primer pairs corresponding to mSSCIR4 and mSSCIR164 loci were used (Raboin *et al.* 2006). Fluorescent labeling and electrophoresis were performed as described for AFLP in the Reunion panel. Whole genome profiling was enriched with DArT markers (Heller-Uszynska *et al.* 2011). Total DNA was sent for genotyping to the commercial company Diversity Arrays Technology Pty Ltd (Yarralumla, Australia). A total of 3337 markers (1408 AFLP, 1899 DArT and 30 SSR) and 3952 markers (2402 AFLP and 1550 DArT) with a frequency ranging between 0.05 and 0.95 were produced, respectively in the Reunion and Guadeloupe panel. 1259 markers were common to the two panels: 1086 DArTs and 173 AFLP. The few number of common AFLP markers is explained by the difficulty to match fingerprint profiles due to the different technologies used to generate the markers for the GUA and RUN panels.

Statistical analysis

Prediction of genotypic value of cultivars

Aphid weekly infestation data from the Reunion trial was characterized by the Area Under Incidence Progression Curve (AUIPC), computed for each of the three cropping seasons separately. AUIPC data were analyzed with a mixed model where cultivar, cultivar x cropping season and block were random effects, while replication and cropping season were fixed effects. To achieve normal distribution of residuals, AUIPC was submitted to a square root transformation prior to analysis. Analysis was performed with SAS procedure Mixed (SAS Institute 2008). We tested two possible improvements of the mixed model used to analyze the AUIPC data: (i) taking into account the correlation of residuals across years due to observations carried out on the same elementary plots (longitudinal data), and (ii) taking into account the spatial correlation among residuals due to aggregative spatial distribution of aphid populations in the field, using methods as described by Littell *et al.* (2006). None of these two modifications proved to be significant, according to a chi-square test performed on the likelihood ratio statistic (Littell *et al.* 2006). The initial mixed model (no longitudinal and no spatial effects) was therefore used to analyze the AUIPC data.

The proportion of stalks infected by SCYLV was analyzed with a generalized mixed model (Logit link function, binomial distribution) using SAS procedure Glimmix and following guidelines from Bolker *et al.* (2008). In Guadeloupe, the model was designed to take into account location and replication as fixed effects, and cultivar and cultivar x location as random effects. In Reunion, replication was a fixed effect, while cultivar and block were random effects. The genotypic value of cultivars was predicted by broad space Best Linear Unbiased Predictors (BLUPs) (McLean *et al.* 1991; Littell *et al.* 2006) computed by adding the solution vector of the random cultivar effect of each model to the trial mean and to the mean of the fixed effects. These values were further used to perform marker-trait association analysis.

The genotypic and environmental correlation coefficients between SCYLV and aphid incidence were computed following Gallais (1990) by performing a multivariate analysis of variance (MANOVA) with the SAS Glim procedure, using regular transformations (square root for AUIPC and $\arcsin(\sqrt{\cdot})$ for SCYLV incidence) to achieve normal distribution of residuals.

Panel genetic structure

Genetic structure of each panel was inferred from DArT markers with a Principal Component Analysis (PCA) approach (Patterson *et al.* 2006), using the SAS Princomp procedure (SAS Institute 2008). Result from PCA may be distorted if uncorrected LD exists between markers (Patterson *et al.* 2006). Therefore, to identify markers that were in LD, we applied the method of Patterson *et al.* (2006) and removed a marker from every pair of markers that were in tight LD. Markers were considered in tight LD when they were significantly associated according to a Fisher's exact test, using the Bonferroni correction for multiple testing. This resulted in the selection of 1096 markers in the Reunion panel and 867 markers in the Guadeloupe panel. The marker matrix was standardized (mean centering and standard deviance scaling) prior to performing PCA. Missing data were replaced by zero value after the standardization step. Four cultivars from the Reunion panel and two from the Guadeloupe panel were identified as outliers in the PCA analysis, i.e. their coordinates in at least one of the ten first PCA axes were at least five fold larger than the within axe standard deviation. These cultivars were therefore removed from the analysis.

Marker-trait associations

Detections of marker-trait associations were carried out with the mixed model approach as described by Yu *et al.* (2006). Analyses were performed with the TASSEL 2.1. software (Bradbury *et al.* 2007) with the GLM model (coordinates along PCA axes used as fixed effect covariates) and with the MLM model (same as GLM plus a random polygenic effect added through a variance-covariance matrix derived from a kinship matrix computed in TASSEL). The number of PCA axes included in the models was determined following the two stage procedure described by Zhu and Yu (2009). In stage 1, the number of significant ($P < 0.05$) principal components was assessed by the Tracy-Widom test as detailed by Patterson *et al.* (2006). In stage 2, we used a linear model to fit each phenotypic trait value with an increasing number of PCA axes coordinate covariates, up to the number of significant dimensions identified in the first stage. The number of dimensions corresponding to the lower Bayesian Information Criterion (BIC) value was chosen.

Markers associated to at least one trait at a marker-wise $P < 0.05$ were grouped in haplotypes, by using the results from pairwise association Fisher's exact tests with Bonferroni

correction, according to the procedure described by Raboin *et al.* (2008). This marker grouping in haplotypes was performed separately in each panel. Final marker-trait associations were considered using (i) marker-wise P value, (ii) P value at a genome-wise (GW) type-1 error rate (based on 1,000 permutation tests). Marker detected at the 0.05 GW type-I error level in a panel and detected at the marker-wise 0.05 type-I error level in the other panel were considered as ‘confirmed linkage’ by analogy with the criterion defined by Lander and Kruglyak (1995) for comparison of marker-trait associations across panels.

Results

SCYLV and aphid incidence

Cultivar mean SCYLV incidence measured by TBIA diagnosis exhibited a U-shaped distribution in Reunion and an L-shaped distribution in Guadeloupe (Fig. 1 and Fig. 2). Nineteen cultivars in Reunion and only one cultivar in Guadeloupe were free of SCYLV. Among the 19 SCYLV-free cultivars from the Reunion panel, only five were present in the Guadeloupe panel: one was free of SCYLV in Guadeloupe, two exhibited a virus incidence lower than 5% contaminated stalks, and two were infected at 33 and 22 %. Comparison of virus incidence between Reunion and Guadeloupe on the 29 shared cultivars with a generalized mixed model using SAS proc Glimmix revealed a significant site x cultivar random interaction (likelihood ratio test: chi-square =148.5, df = 1, $P < 0.0001$). Some cultivars appeared more infected in Guadeloupe than in Reunion, while the reverse was observed for some cultivars (Supplementary Figure 1).

The viral genotypes BRA-PER and REU were detected in both cultivar panels, while CUB was detected only in Guadeloupe (Table 1). Prevalence of viral genotypes varied greatly between the two panels. In Reunion, the REU viral genotype was found in 96 % of the infected cultivars and the BRA-PER genotype only in 8%. In Guadeloupe CUB, REU, and BRA-PER genotypes were found in 98%, 72% and 62% of the infected cultivars respectively.

The distribution of aphid incidence means among the 184 cultivars exhibited a bell shaped distribution with a slight right dissymmetry due to few highly susceptible cultivars (Fig. 3). Nineteen resistant cultivars exhibited an aphid incidence mean significantly ($P < 0.05$) lower than the trial mean, 13 susceptible cultivars exhibited an aphid incidence mean significantly

($P < 0.05$) higher than the trial mean, and the remaining 151 intermediate cultivars exhibited an aphid incidence mean not significantly different from the trial mean ($P > 0.05$).

Phenotypic correlation between SCYLV incidence and aphid infestation was positive and significant ($R = 0.108$; $P = 0.0093$). Genotypic correlation coefficient between both traits was low (0.162), whereas the environmental correlation was close to zero (0.010). Comparison of virus incidence mean between the three aphid susceptibility classes (resistant, intermediate, susceptible), using SAS proc Glimmix, revealed a highly significant class effect ($F = 83.03$; $df = 2, 196$; $P < 0.0001$). The aphid resistant class had a lower mean (\pm SE) virus incidence (47.4 ± 2.1 %) than the susceptible (71.8 ± 2.1 %) and the intermediate (74.1 ± 0.6 %) classes. Virus incidences in susceptible and intermediate classes were not significantly different ($P > 0.05$).

Panel structure

In Reunion, the analysis of panel structure with PCA yielded 24 axes that were significant ($P < 0.05$) by the Tracy-Widom test, whereas in Guadeloupe, the number of significant PCA axes was 15. Fitting of each phenotypic trait value with an increasing number of significant dimensions showed that minimum BIC was achieved with five dimensions for the virus incidence in the Reunion panel and with one dimension in the Guadeloupe panel. For aphid incidence, the BIC was minimal with zero dimension.

Therefore, axes 1 to 5 were used as covariates in the GLM and MLM models for virus incidence trait in Reunion, whereas only the first axis was used in Guadeloupe. No correction for structure was applied for the aphid incidence trait. Structure covariates had a highly significant effect on virus incidence traits, i.e. $P < 0.0001$ ($R^2 = 0.29$) for five axes in Reunion, and $P < 0.0001$ ($R^2 = 0.10$) for one axis in Guadeloupe.

Marker trait-associations

Structure had a strong effect on the detection of marker-trait associations. For marker-virus incidence association, 458 significant markers were detected in Reunion at the marker-wise $\alpha = 0.05$ type-1 error level without taking the structure into account, versus 214

markers at the same error level taking structure into account. 545 significant markers were detected in Guadeloupe without structure and 334 markers with structure. Analysis of marker-trait associations with the GLM model, using 1,000 permutations for the control of GW type-1 error for virus incidence, revealed 21 significant markers in Reunion, defining 12 independent haplotypes (Table 2, Supplementary Table 1). Seven of these haplotypes were associated with resistance and five with susceptibility. In Guadeloupe (Table 2, Supplementary Table 2), 32 significant associations were detected, corresponding to 19 haplotypes. Sixteen of these haplotypes were associated with resistance and three with susceptibility. A total of 30 haplotypes were significantly linked ($P < 0.05$ with GW type-1 error control) to virus incidence in at least one of the two panels (Table 2). Analysis of marker-trait associations with MLM model yielded probability values close to values obtained with GLM model (Supplementary Table 1 and 2). Haplotypes associated with SCYLV incidence explained 4 to 8% and 5 to 11% of the phenotypic variation (R^2) in the Reunion (Supplementary Table 1) and Guadeloupe (Supplementary Table 2) panels respectively. Haplotypes significantly associated with aphid infestation explained 5 to 9% of the phenotypic variation (Supplementary Table 3).

Among the markers that were significantly associated to virus incidence either in Reunion or in Guadeloupe for virus incidence, 28 were coded in both panels, defining 17 independent common haplotypes. Five of these common haplotypes were confirmed linkages (Table 2): (i) one haplotype (R-1-19_G-1-11) was significant simultaneously in both panels at the 0.05 GW type-I error level, (ii) four haplotypes (R-1-13_G-1-66, R-1-60_G-1-48, R-1-42_G-1-34 and R-1-36_G-1-25) were significant at the 0.05 GW type-I error level in one panel and significant at the marker-wise 0.05 type-I error level in the second panel.

Analysis of marker-trait associations (GLM model) for resistance to aphids detected 25 significant markers, defining 18 independent haplotypes (Table 2, Supplementary Table 3). Among these 18 haplotypes, 13 were associated with resistance and five with susceptibility. Three haplotypes (R-1-41_G-1-68, R-1-16_G-1-22 and R-1-37_G-1-25) were associated with both aphid and virus incidence, with the same qualitative effect (resistance or susceptibility) for both traits (Table 2).

Discussion

In our study, we identified 30 haplotypes significantly associated with SCYLV incidence on a basis of a 0.05 GW type-I error threshold either in Reunion (12) or Guadeloupe (19) panels, after controlling the population structure effect. Twenty-two of these haplotypes were associated with resistance and eight with susceptibility. Interestingly, five of these haplotypes were associated with SCYLV incidence in both panels/environments. This study is the first report for sugarcane of a marker-trait association based on a cross validation between two independent association mapping studies.

Based on the serological TBIA indexations (testing for the incidence of any virus genotypes), we obtained large agronomic segregations in both panels that were well adapted to mapping association studies for virus resistance. The only noticeable difference between the two TBIA-based segregations concerned the number of SCYLV free cultivars (19 cultivars in Reunion and only one in Guadeloupe). This difference is mainly due to the choice of cultivars, since only five of the 19 SCYLV free cultivars present in Reunion panel were also present in the Guadeloupe panel. Nevertheless, a significant cultivar x site interaction was demonstrated on the 29 common cultivars. This finding could be related to differences in virus genotype prevalence between the two locations, combined with the differences in infection capacity between the three viral genotypes, as demonstrated by Abu Ahmad *et al.* (2007). Considering the markedly distinct epidemiological conditions prevailing in the two experimental regions, the five haplotypes co-detected in Reunion and Guadeloupe should harbor alleles determining resistance effects of a broad spectrum nature, rather than specific to a particular viral genotype.

The study of resistance to aphids in Reunion allowed the detection of 25 significant markers, grouped into 18 independent haplotypes. Thirteen of these haplotypes had a positive effect on resistance and five haplotypes had a negative effect on it. A significant, although modest, positive phenotypic correlation, resulting mainly from genetic correlation was revealed between aphid incidence and SCYLV incidence, as cultivars resistant to *M. sacchari* showed significantly lower levels of virus incidence. Three explanations could account for this finding. First, clusters containing distinct genes that control virus and vector resistance may exist. Such clusters of resistance genes, containing insect and disease resistance loci have been described in maize (McMullen and Simcox 1995). Second, resistance to aphids may be efficient enough to reduce SCYLV transmission. Some authors (van Emden 2007) have

claimed the usefulness of breeding for vector resistance to control infection by semi-persistently or persistently transmitted viruses. This has been reported previously on *Triticum monococcum* TM 44 line resistant to *Sitobion avenae* showing reduced BYDV-PAV transmission (Tanguy and Dedryver 2009). Delatte *et al.* (2006) showed that resistance to the whitefly vector *Bemisia tabaci* allowed a quantitative resistance to the *Tomato yellow leaf curl virus* in *Lycopersicon pimpinellifolium*. Similarly, Dintinger *et al.* (2005) demonstrated that resistance to the planthopper vector *Peregrinus maidis* in some maize lines was associated with resistance to *Maize stripe virus* and *Maize mosaic virus*. A third explanation could be an increased aphid preference for virus infected plants. This has already been described with the *Potato leafroll virus* (Luteoviridae: *Polerovirus*) in potato, which induces emission of volatiles by virus infected leaves acting as attractant to the aphid *Myzus persicae* (Alvarez *et al.* 2007).

Haplotypes associated with SCYLV or aphid incidence explained 4 to 11% of the phenotypic variation. Considering the number of accessions used, these values have the same order of magnitude as those found in similar association mapping studies for traits of interest in other crops, such as sorghum (de Alencar Figueiredo *et al.* 2010) or cotton (Wegrzyn *et al.* 2010). Our results confirm the potential of association mapping in sugarcane, and identify five haplotypes potentially useful for Yellow Leaf resistance breeding.

Ethical standards

The authors declare that the experiments presented in this publication comply with the current laws of France.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgments

This work was funded by the Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), by the Conseil Régional de la Réunion and

by the European Union (European regional development fund - ERDF), by ANR Délicas project grant ANR-08-GENM-001. The authors wish to thank A. Anglade, A. Avaby, T. Atiama, C. Hatik, C. Lallemand, M. Payet, L. Poncharville, I. Promi, H. Telismart, R. Tibère, and L. Touby for their technical assistance. We thank L. Camus-Kulandaivelu for critical review of the manuscript.

References

- Abu Ahmad Y, Costet L, Daugrois J-H, Nibouche S, Letourmy P, Girard J-C, Rott P (2007) Variation in infection capacity and in virulence exists between genotypes of *Sugarcane yellow leaf virus*. *Plant Dis* 91:253-259
- Abu Ahmad Y, Rassaby L, Royer M, Borg Z, Braithwaite KS, Mirkov TE, Irey MS, Perrier X, Smith GR, Rott P (2006a) Yellow leaf of sugarcane is caused by at least three different genotypes of *Sugarcane yellow leaf virus*, one of which predominates on the Island of Reunion. *Arch of Virol* 151:1355-1371
- Abu Ahmad Y, Royer M, Daugrois J-H, Costet L, Lett J-M, Victoria JJ, Girard J-C, P. R (2006b) Geographical distribution of four *Sugarcane yellow leaf virus* genotypes. *Plant Dis* 90:1156-1160
- Akbar W, Showler AT, Reagan TE, White WH (2010) Categorizing sugarcane cultivar resistance to the sugarcane aphid and yellow sugarcane aphid (Hemiptera: *Aphididae*). *J Econ Entomol* 103:1431-1437
- Alvarez AE, Garzo E, Verbeek M, Vosman B, Dicke M, Tjallingii WF (2007) Infection of potato plants with *Potato leafroll virus* changes attraction and feeding behaviour of *Myzus persicae*. *Ent Exp et Applic* 125:135-144
- Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, White J-SS (2008) Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol Evol* 24 127-135
- Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633-2635
- Butterfield M (2007) Marker assisted breeding in sugarcane: a complex polyploid. PhD thesis University of Stellenbosch
- Comstock JC, Miller JD, Schnell RJ (2001) Incidence of *Sugarcane yellow leaf virus* in clones maintained in the world collection of sugarcane and related grasses at the United States National Repository in Miami, Florida. *Sugar-Tech* 3:128-133
- D'Arcy C, Domier L (2005) Virus taxonomy. VIIIth Report of the International Committee on Taxonomy of Viruses. In: Fauquet C, Mayo M, Maniloff J, Desselberger U, Ball L (eds) *Luteoviridae*. Elsevier Academic Press, New York, pp 891-900
- de Alencar Figueiredo L, Sine B, Chanterreau J, Mestres C, Flidel G, Rami JF, Glaszmann JC, Deu M, Courtois B (2010) Variability of grain quality in sorghum: association with polymorphism in *Sh2*, *Bt2*, *SssI*, *Ae1*, *Wx* and *O2*. *Theor Appl Genet* 121:1171-1185
- Daugrois J-H., Edon-Jock C., Bonoto S., Vaillant J., Rott P. 2011. Spread of *Sugarcane yellow leaf virus* in initially disease-free sugarcane is linked to rainfall and host resistance in the humid tropical environment of Guadeloupe. *Eur J Plant Pathol* 129:71–80

- Delatte H, Holota H, Reynaud B, Dintinger J (2006) Characterisation of a quantitative resistance to vector transmission of *Tomato yellow leaf curl virus* in *Lycopersicon pimpinellifolium*. *Eur J Plant Pathol* 114:245-253
- Dintinger J, Boissot N, Chiroleu F, Hamon P, Reynaud B (2005) Evaluation of maize inbreds for *Maize stripe virus* and *Maize mosaic virus* resistance: disease progress in relation to time and the cumulative number of planthoppers. *Phytopathology* 95:600-607
- Fartek B (2011) Résistance de la canne à sucre à la maladie de la feuille jaune et à son vecteur *Melanaphis sacchari* : Association marqueur-trait dans une population de cultivars, mécanismes de résistance au vecteur, et diversité génétique du vecteur. PhD thesis, Université de la Réunion, Saint-Pierre
- Gallais A (1990) Théorie de la sélection en amélioration des plantes. Collection sciences agronomiques
- Heller-Uszynska K, Uszynski G, Huttner E, Evers M, Carlig J, Caig V, Aitken K, Jackson P, Piperidis G, Cox M, Gilmour R, D'Hont A, Butterfield M, Glaszmann J-C, Kilian A (2011) Diversity Arrays Technology effectively reveals DNA polymorphism in a large and complex genome of sugarcane. *Mol Breed* 28:37-55
- Hoarau JY, Offmann B, D'Hont A, Risterucci AM, Roques D, Glaszmann JC, Grivet L (2001) Genetic dissection of a modern sugarcane cultivar (*Saccharum* spp.). I. Genome mapping with AFLP markers. *Theor Appl Genet* 103:84-97
- Jannoo N, Grivet L, Dookun A, D'Hont A, Glaszmann JC (1999) Linkage disequilibrium among modern sugarcane cultivars. *Theor Appl Genet* 99:1053-1060
- Komor E (2011) Susceptibility of sugarcane, plantation weeds and grain cereals to infection by *Sugarcane yellow leaf virus* and selection by sugarcane breeding in Hawaii. *Eur J Plant Pathol* 129:379-388
- Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 11:241-247
- Littell RC, Milliken GA, Stroup WW, Wolfinger RD, Schabenberger O (2006) SAS system for mixed models, second edition. SAS Institute Inc., Cary, NC
- McLean RA, Sanders WL, Stroup WW (1991) A Unified Approach to Mixed Linear Models. *Am Stat* 45:54-64
- McMullen M, Simcox K (1995) Genomic organization of disease and insect resistance genes in maize. *Mol Plant Microbe In* 8:811-815
- Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genet* 2:2074-2092
- Raboin L-M, Pauquet J, Butterfield M, D'Hont A, Glaszmann J-C (2008) Analysis of genome-wide linkage disequilibrium in the highly polyploid sugarcane. *Theor Appl Genet* 116:701-714
- Raboin L, Oliveira K, Lecunff L, Telismart H, Roques D, Butterfield M, Hoarau J, D'Hont A (2006) Genetic mapping in sugarcane, a high polyploid, using bi-parental progeny: identification of a gene controlling stalk colour and a new rust resistance gene. *Theor Appl Genet* 112:1382-1391
- Raboin LM (2005) Génétique de la résistance au charbon de la canne à sucre causé par *Ustilago scitaminea* : caractérisation de la diversité génétique du pathogène, cartographie de QTL dans un croisement bi-parental et étude d'associations dans une population de cultivars modernes. Doctorat thesis, ENSAM, Montpellier
- Rassaby L, Girard JC, Lemaire O, Costet L, Irey MS, Kodja H, Lockhart BEL, Rott P (2004) *Spread of Sugarcane yellow leaf virus* in sugarcane plants and fields on the island of Reunion. *Plant Pathol* 53:117-125

- Rassaby L, Girard JC, Letourmy P, Chaume J, Irely MS, Lockhart BEL, Kodja H, Rott P (2003) Impact of *Sugarcane yellow leaf virus* on sugarcane yield and juice quality in Réunion Island. *Eur J Plant Pathol* 109:459-466
- Rott P, Mirkov TE, Schenck S, Girard JC (2008) Recent advances in research on *Sugarcane yellow leaf virus*, the causal agent of sugarcane yellow leaf. *Sugar Cane Int* 26:18-27
- SAS Institute (2008) SAS OnlineDoc® 9.2. SAS Institute Inc., Cary, NC, USA
- Schenck S (1990) Yellow leaf syndrome—a new sugarcane disease. Hawaiian Sugar Planters Association: Annual Report 38
- Schenck S (1997) Advances in control of yellow leaf syndrome. Pathology Report 67. Hawaii Agriculture Research Center, 4p
- Schenck S, Hu JS, Lockhart BE (1997) Use of a tissue blot immunoassay to determine the distribution of *Sugarcane yellow leaf virus* in Hawaii. *Sugar Cane* 4:5-8
- Schenck S and Lehrer AT (2000) Factors affecting the transmission and spread of *Sugarcane yellow leaf virus*. *Plant Dis* 84:1085-1088
- Tanguy S, Dedryver CA (2009) Reduced BYDV–PAV transmission by the grain aphid in a *Triticum monococcum* line. *Eur J Plant Pathol* 123:281-289
- van Emden H (2007) Host-plant resistance. In: van Emden H, Harrington R (eds) *Aphids as crop pests*. CAB International, Wallingford, UK, pp 447-468
- Vega J, Scagliusi SMM, Ulian EC (1997) Sugarcane yellow leaf disease in Brazil: Evidence of association with a luteovirus. *Plant Dis* 81:21-26
- Viswanathan R, Balamuralikrishnan M, Karuppaiah R (2008) Identification of three genotypes of *Sugarcane yellow leaf virus* causing yellow leaf disease from India and their molecular characterization. *Virus Genes* 37:368-379
- Wang MQ and Zhou GH (2010) A near-complete genome sequence of a distinct isolate of *Sugarcane yellow leaf virus* from China, representing a sixth new genotype. *Virus Genes* 41:268-272
- Wegrzyn JL, Eckert AJ, Choi M, Lee JM, Stanton BJ, Sykes R, Davis MF, Tsai C-J, Neale DB (2010) Association genetics of traits controlling lignin and cellulose biosynthesis in black cottonwood (*Populus trichocarpa*, Salicaceae) secondary xylem. *New Phytol* 188:515-532
- Wei X, Jackson P, McIntyre C, Aitken K, Croft B (2006) Associations between DNA markers and resistance to diseases in sugarcane and effects of population substructure. *Theor Appl Genet* 114:155-164
- Wei X, Jackson PA, Hermann S, Kilian A, Heller-Uszynska K, Deomano E (2010) Simultaneously accounting for population structure, genotype by environment interaction, and spatial variation in marker-trait associations in sugarcane. *Genome* 53:973-981
- Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* 38:203-208
- Zhu C and Yu J (2009) Nonmetric multidimensional scaling corrects for population structure in association mapping with different sample types. *Genetics* 182:875-888

Tables

Table 1 Estimation of prevalence of viral genotypes in each panel: frequency of cultivars infected by each viral genotype among infected cultivars

Viral genotype	Guadeloupe	Reunion
REU	0.72	0.96
BRA-PER	0.62	0.08
CUB	0.98	0

Table 2 List of haplotypes significantly associated with resistance to SCYLV in Reunion (RUN) and Guadeloupe (GUA) panels and associated with resistance to *M. sacchari* in the Reunion panel, using correction for panel structure (GLM model)

haplotypes		Virus incidence			Effect ^c	Aphid incidence RUN	
		genome-wise significance ^a		marker-wise significance in RUN or GUA ^b		Genome-wise significance ^a	Effect ^c
RUN	GUA	RUN	GUA	RUN or GUA ^b			
R-1-19	G-1-11	***	*	***	R	NS	-
R-1-13	G-1-66	***	NS	**	S	***	S
R-1-60	G-1-48	NS	***	*	R	NS	-
R-1-42	G-1-34	NS	***	*	R	NS	-
R-1-36	G-1-25	*	NS	*	R	*	R
R-1-79	G-1-19	***	NS	NS	R	NS	-
R-1-16	G-1-1	**	NS	NS	S	*	S
R-1-172	-	***	-	-	R	NS	-
R-1-133	-	*	-	-	R	NS	-
R-1-160	G-1-95	*	NS	NS	S	NS	-
R-1-154	-	***	-	-	R	NS	-
R-1-182	G-1-142	*	NS	NS	R	NS	-
R-1-161	-	*	-	-	S	NS	-
R-1-135	G-1-165	***	NS	NS	S	NS	-
-	G-1-147	-	***	-	R	-	-
-	G-1-168	-	*	-	R	-	-
-	G-1-166	-	*	-	R	-	-
-	G-1-167	-	***	-	R	-	-
R-1-165	G-1-156	NS	***	NS	S	NS	-
-	G-1-148	-	***	-	R	-	-
-	G-1-129	-	***	-	R	-	-
R-1-183	G-1-10	NS	***	NS	R	NS	-
R-1-115	G-1-15	NS	***	NS	R	NS	-
-	G-1-20	-	***	-	R	-	-
-	G-1-26	-	***	-	S	-	-
R-1-174	G-1-37	NS	***	NS	R	NS	-
-	G-1-40	-	***	-	R	-	-
R-1-38	G-1-41	NS	***	NS	S	NS	-
R-1-64	G-1-52	NS	*	NS	R	NS	-
R-1-86	G-1-70	NS	***	NS	R	NS	-
R-1-180	-	NS	-	-	-	*	R
R-1-185	-	NS	-	-	-	*	R
R-1-125	G-1-88	NS	NS	-	-	*	R
R-1-122	-	NS	-	-	-	*	R
R-1-127	G-1-154	NS	NS	-	-	***	R
R-1-10	G-1-92	NS	NS	-	-	***	R
R-1-24	G-1-158	NS	NS	-	-	***	S
R-1-27	-	NS	-	-	-	***	R
R-1-29	G-1-39	NS	NS	-	-	***	R
R-1-30	G-1-175	NS	NS	-	-	***	S
R-1-41	G-1-46	NS	NS	-	-	***	R
R-1-44	G-1-90	NS	NS	-	-	***	S
R-1-61	G-1-111	NS	NS	-	-	*	R
R-1-62	G-1-50	NS	NS	-	-	*	R
R-1-75	G-1-61	NS	NS	-	-	***	R

^a highest significance level for marker-trait associations within haplotype, using a permutation test to control genome-wise type-1 error: $P < 0.001$ (***), $P < 0.01$ (**), $P < 0.05$ (*), $P > 0.05$ (NS). - : haplotype not genotyped in the considered panel

^b confirmation of associations for virus incidence: highest marker-wise significance level within haplotype in one panel (** $P < 0.01$; * $P \leq 0.05$; NS: $P > 0.05$) for haplotypes containing markers significant at the genome-wise level in the other panel. - : the haplotype is genotyped in one panel only

^c haplotype effect : R = resistance, S = susceptibility, - : haplotype not significantly associated with resistance

Figure captions

Fig. 1 Frequency distribution of the mean percentage of stalks infected by SCYLV among 184 cultivars in Reunion (cultivar phenotypic mean over three replications)

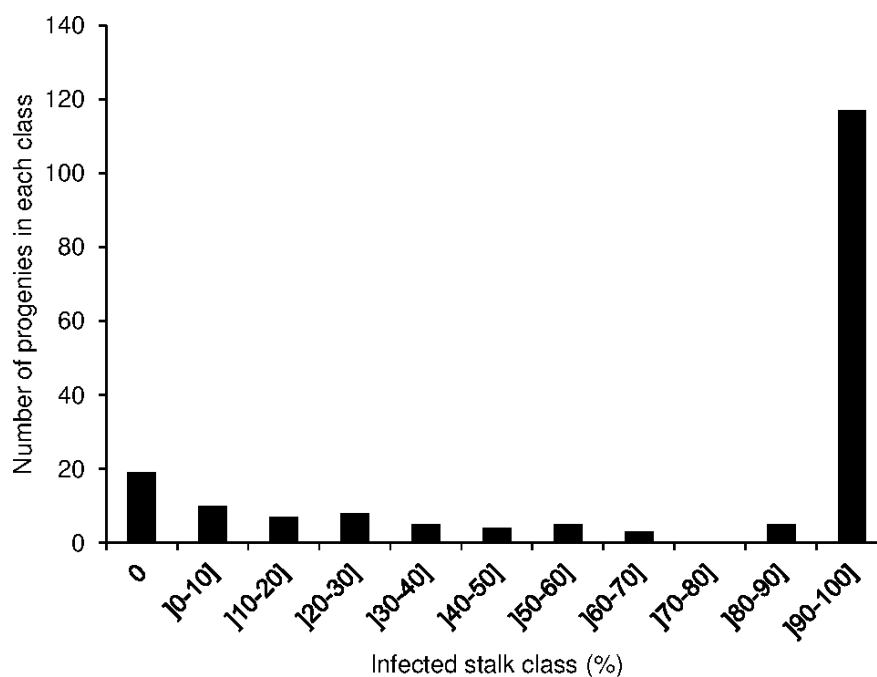


Fig. 2 Frequency distribution of the mean percentage of stalks infected by SCYLV among 189 cultivars in Guadeloupe (cultivar phenotypic mean over two trials, three replications each)

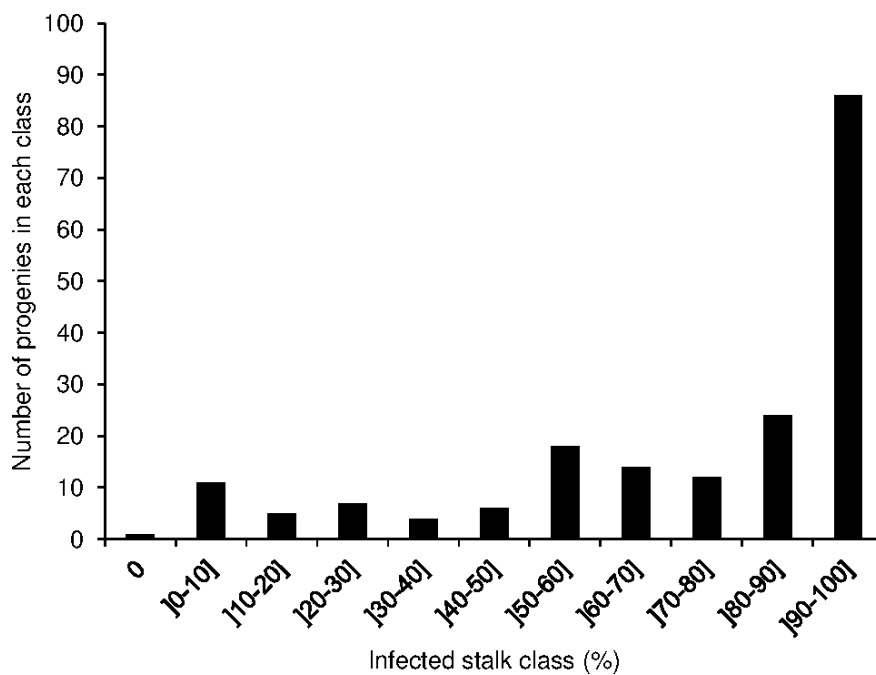
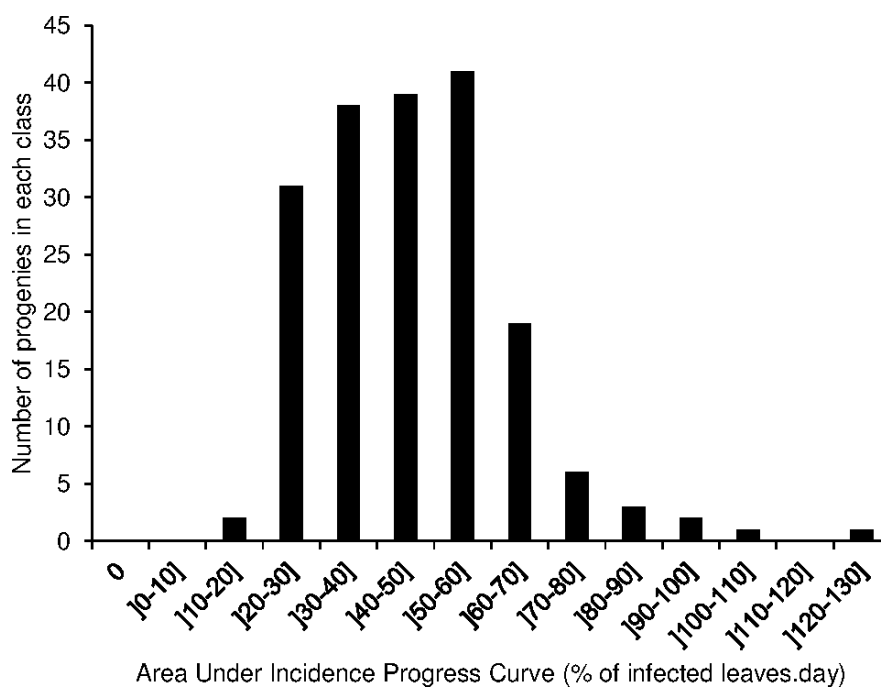
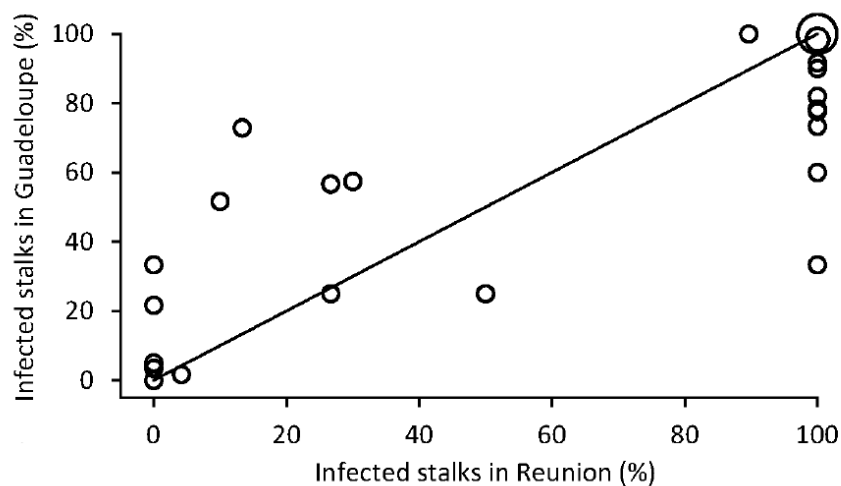


Fig. 3 Frequency distribution of the mean aphid incidence among 184 cultivars in Reunion (phenotypic mean over three replications, during three years and 29 observations)



Supplementary tables



Supplementary table 1: Comparison of the mean percentage of stalks infected by SCYLV among 29 cultivars common to the Guadeloupe and Reunion panels. Line represents the median. Surface of dots is proportional to the number of cultivars (i.e. 1, 2 or 6 cultivars from smaller to larger).

Supplementary table 2: List of markers significantly associated with resistance to SCYLV in Reunion panel, using correction for panel structure (GLM model) or corrections for structure and kinship (MLM model).

haplotype	marker ^a	number of genotyped cultivars	marker frequency	GLM			MLM marker-wise <i>P</i> value	GLM marker-wise <i>P</i> value in Guadeloupe ^c
				partial coefficient of determination <i>R</i> ²	marker-wise <i>P</i> value	genome-wise significance ^b		
R-1-172	R-A-540	183	0.58	0.08	4.4E-06	***	5.6E-06	-
R-1-133	R-A-154	184	0.11	0.04	1.9E-03	*	1.8E-03	-
R-1-160	D-186	147	0.89	0.05	2.1E-03	*	2.5E-03	6.0E-01
R-1-154	D-1820	151	0.07	0.06	6.3E-04	***	6.4E-04	-
R-1-182	D-1326	170	0.29	0.04	3.2E-03	*	1.9E-03	5.5E-01
R-1-161	D-556	176	0.90	0.04	1.7E-03	*	2.2E-03	-
R-1-135	D-1712	177	0.94	0.05	8.8E-04	***	1.2E-03	8.8E-01
R-1-16	R-A-65	167	0.46	0.05	1.0E-03	*	2.8E-03	3.6E-01
R-1-16	R-A-863	182	0.39	0.04	1.4E-03	*	1.9E-03	-
R-1-16	D-1433	175	0.57	0.04	2.8E-03	*	3.4E-03	-
R-1-16	D-1474	177	0.61	0.04	2.0E-03	*	2.0E-03	-
R-1-16	D-1275	181	0.58	0.04	1.4E-03	**	1.4E-03	-
R-1-16	D-1405	180	0.56	0.04	1.9E-03	*	1.9E-03	-
R-1-19	R-A-903	183	0.30	0.04	1.1E-03	**	1.7E-03	-
R-1-19	D-52	172	0.18	0.06	1.2E-04	***	1.8E-04	<u>3.3E-03</u>
R-1-19	D-817	183	0.11	0.05	3.4E-04	***	4.7E-04	<u>1.8E-03</u>
R-1-19	D-212	184	0.11	0.06	2.0E-04	***	3.0E-04	<u>1.8E-03</u>
R-1-36	D-371	182	0.12	0.04	2.3E-03	*	3.5E-03	<u>1.4E-02</u>
R-1-13	D-757	180	0.85	0.04	1.3E-03	*	1.8E-03	9.0E-02
R-1-13	D-1342	181	0.87	0.04	8.7E-04	***	1.4E-03	<u>7.2E-03</u>
R-1-79	D-1220	175	0.18	0.05	6.8E-04	***	1.3E-03	2.3E-01

^a marker names beginning by R-A: AFLP; by D: DArTs

^b significant association at $P < 0.001$ (***), $P < 0.01$ (**), $P < 0.05$ (*), using a permutation test to adjust for genome-wise type-1 error.

^c underlined values correspond to confirmed linkages, i.e. linkage that are significant in Reunion with genome-wise type-1 error control and significant at a marker-wise $P < 0.05$ in Guadeloupe. Markers not coded in Guadeloupe panel are signaled by '-'.

Supplementary table 3: List of markers significantly associated with resistance to SCYLV in Guadeloupe panel, using correction for panel structure (GLM model) or corrections for structure and kinship (MLM model).

haplotype	marker ^a	number of genotyped cultivars	marker frequency	GLM			MLM marker-wise <i>P</i> value	GLM marker-wise <i>P</i> value in Reunion ^c
				partial coefficient of determination R ²	marker-wise <i>P</i> value	genome-wise significance ^b		
G-1-147	G-A-1137	187	0.16	0.06	5.5E-04	***	1.2E-03	-
G-1-168	G-A-1266	176	0.32	0.05	1.2E-03	*	2.2E-03	-
G-1-166	G-A-401	115	0.12	0.07	1.9E-03	*	5.2E-03	-
G-1-167	G-A-597	174	0.20	0.08	1.2E-04	***	4.1E-04	-
G-1-156	G-A-1728	174	0.63	0.07	2.6E-04	***	9.0E-04	8.7E-01
G-1-148	G-A-98	186	0.18	0.08	3.0E-05	***	6.1E-05	-
G-1-129	G-A-232	186	0.11	0.06	3.2E-04	***	2.4E-03	-
G-1-10	D-228	187	0.13	0.06	3.3E-04	***	6.6E-05	7.1E-02
G-1-11	D-817	188	0.09	0.05	1.8E-03	*	7.0E-03	<u>4.5E-04</u>
G-1-11	D-212	188	0.09	0.05	1.8E-03	*	7.0E-03	<u>2.9E-04</u>
G-1-15	G-A-1093	188	0.15	0.05	8.8E-04	***	1.4E-02	-
G-1-15	G-A-693	133	0.29	0.08	4.0E-04	***	5.8E-04	-
G-1-20	G-A-1873	161	0.44	0.06	1.4E-03	*	9.2E-03	-
G-1-20	G-A-386	163	0.39	0.07	2.6E-04	***	5.8E-04	-
G-1-26	G-A-1758	165	0.23	0.06	9.2E-04	***	4.0E-03	-
G-1-26	G-A-472	168	0.35	0.05	1.5E-03	*	6.6E-03	-
G-1-34	G-A-1776	187	0.14	0.06	4.1E-04	***	1.5E-03	-
G-1-34	D-340	186	0.20	0.11	2.1E-06	***	1.1E-05	<u>4.1E-02</u>
G-1-34	D-1363	188	0.07	0.05	1.8E-03	*	1.3E-02	-
G-1-37	G-A-513	178	0.11	0.07	1.2E-04	***	8.3E-04	-
G-1-37	D-310	187	0.14	0.07	8.6E-05	***	6.5E-04	1.2E-01
G-1-37	D-993	188	0.52	0.06	7.2E-04	***	2.6E-03	-
G-1-40	G-A-1165	183	0.06	0.08	7.9E-05	***	1.3E-04	-
G-1-41	D-1720	188	0.61	0.07	1.2E-04	***	7.2E-04	-
G-1-48	D-102	172	0.65	0.07	3.0E-04	***	1.2E-03	<u>3.1E-02</u>
G-1-48	D-1737	178	0.62	0.05	2.1E-03	*	5.3E-03	<u>4.2E-02</u>
G-1-52	D-299	187	0.07	0.05	1.8E-03	*	3.2E-03	1.2E-01
G-1-52	D-398	187	0.07	0.05	1.8E-03	*	3.2E-03	1.2E-01
G-1-52	D-1757	184	0.08	0.05	1.8E-03	*	2.9E-03	1.2E-01
G-1-70	D-1814	188	0.16	0.05	1.9E-03	*	6.8E-03	7.3E-01
G-1-70	D-1056	187	0.14	0.05	8.1E-04	***	4.3E-03	7.6E-01
G-1-70	D-190	188	0.15	0.05	2.0E-03	*	8.4E-03	7.9E-01

^a Marker names beginning by G-A: AFLP; by D: DArTs

^b significant association at $P < 0.001$ (***), $P < 0.01$ (**), $P < 0.05$ (*), using a permutation test to adjust for genome-wise type-I error.

^c Underlined values correspond to confirmed linkages, i.e. linkage that are significant in Guadeloupe with genome-wise type-I error control and significant at a marker-wise $P < 0.05$ in Reunion. Markers not coded in Reunion panel are signaled by ‘-’.

Supplementary table 4: List of markers significantly associated with resistance to *M. sacchari* in the Reunion panel, using correction for panel structure (GLM model) or corrections for structure and kinship (MLM model).

haplotype	marker ^a	number of genotyped cultivars	marker frequency	GLM			MLM marker- wise <i>P</i> value
				partial coefficient of determination n R ²	marker- wise <i>P</i> value	genome-wise significance ^b	
R-1-180	R-A-371	176	0.26	0.06	1.0E-03	*	4.6E-03
R-1-185	D-945	180	0.84	0.06	1.6E-03	*	7.3E-03
R-1-125	D-921	179	0.08	0.05	2.3E-03	*	1.8E-02
R-1-122	D-913	154	0.86	0.06	1.8E-03	*	7.4E-03
R-1-127	D-735	183	0.89	0.07	5.5E-04	***	2.0E-03
R-1-10	R-A-209	176	0.36	0.07	6.4E-04	***	9.8E-03
R-1-16	D-1038	125	0.38	0.07	2.5E-03	*	4.2E-02
R-1-24	D-1072	173	0.29	0.08	2.8E-04	***	1.6E-03
R-1-27	D-487	182	0.18	0.08	1.3E-04	***	6.3E-04
R-1-27	D-1131	181	0.17	0.09	7.5E-05	***	3.8E-04
R-1-29	R-A-503	181	0.69	0.09	7.2E-05	***	4.6E-04
R-1-29	R-A-868	182	0.60	0.07	3.8E-04	***	2.4E-03
R-1-29	D-831	175	0.79	0.06	1.8E-03	*	3.4E-03
R-1-30	R-A-602	184	0.43	0.07	4.8E-04	***	3.6E-03
R-1-36	D-795	182	0.12	0.06	1.3E-03	*	6.6E-03
R-1-13	R-A-1260	167	0.37	0.07	8.8E-04	***	7.2E-03
R-1-13	D-245	175	0.52	0.09	6.5E-05	***	1.8E-04
R-1-41	D-675	162	0.72	0.06	1.3E-03	*	5.2E-03
R-1-41	D-445	167	0.73	0.08	3.8E-04	***	1.1E-03
R-1-44	D-475	180	0.47	0.07	6.2E-04	***	6.4E-03
R-1-44	D-1741	183	0.45	0.07	4.6E-04	***	4.8E-03
R-1-61	D-1813	180	0.72	0.05	2.1E-03	*	1.5E-02
R-1-62	D-581	180	0.11	0.06	1.6E-03	*	1.2E-03
R-1-75	D-1796	179	0.85	0.08	1.2E-04	***	4.0E-04
R-1-75	D-1582	180	0.86	0.07	2.4E-04	***	8.5E-04

^a marker names beginning by R-A: AFLP; by D: DArTs

^b significant association at $P < 0.001$ (***), $P < 0.01$ (**), $P < 0.05$ (*), using a permutation test to adjust for genome-wise type-I error

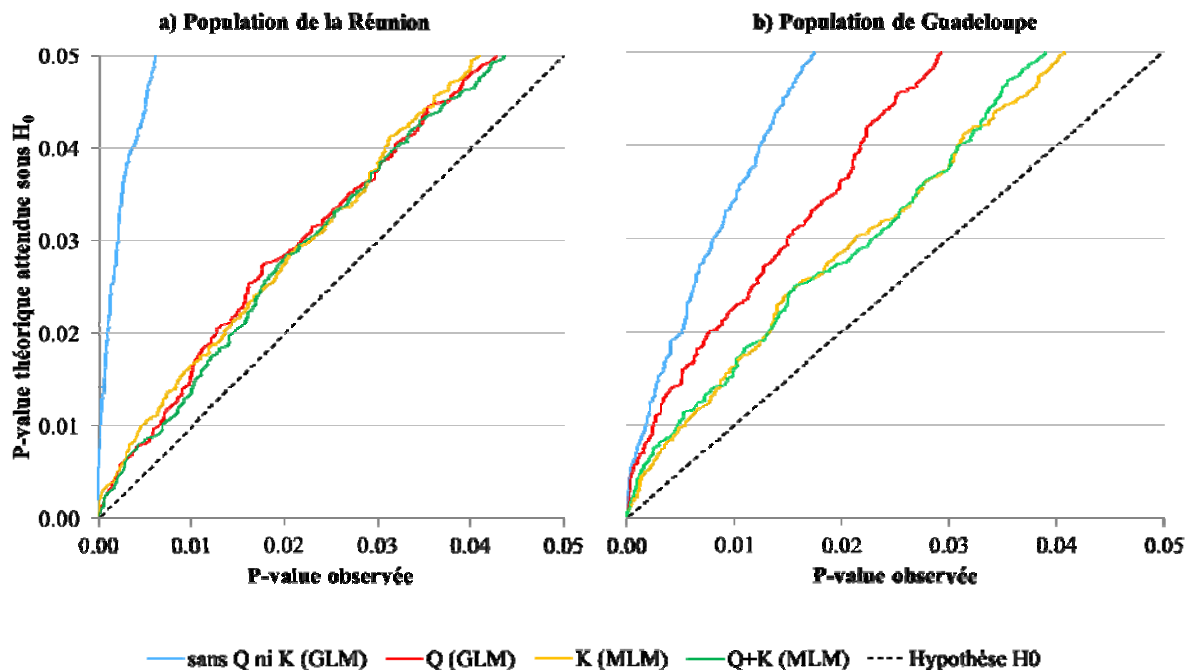


Figure 21 : Distribution cumulée des p-value observées (P_{obs}) par rapport à une loi de probabilité uniforme, des différents modèles d'associations utilisés (Naïf, Q, K, Q+K), concernant la résistance à la feuille jaune. Sous l'hypothèse d'une absence de marqueurs liés au polymorphisme contrôlant la résistance au SCYLV (H_0), le modèle d'analyse qui contrôle le mieux les erreurs de type I tend vers une distribution uniforme des P-values.

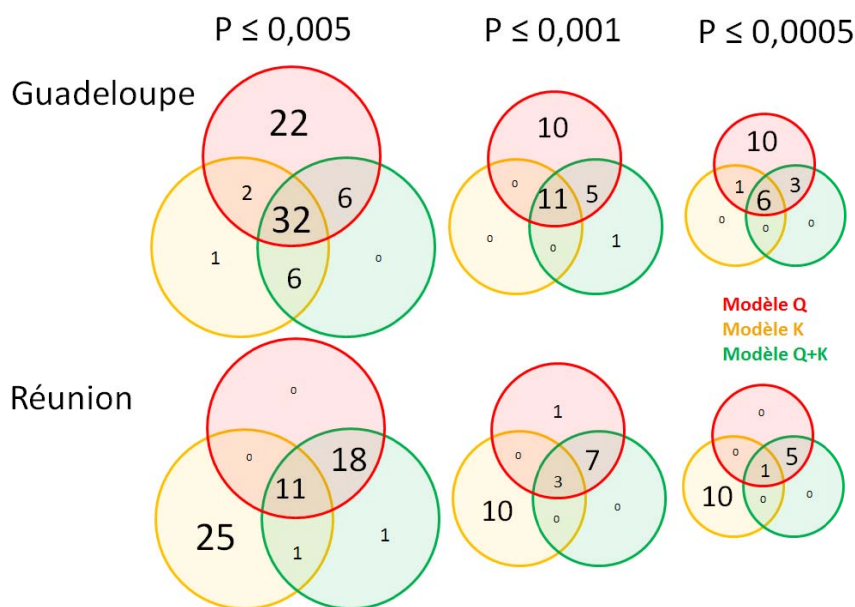


Figure 22: Nombre d'associations détectées par chacun des modèles (Q, K et Q+K) pour la résistance au SCYLV, dans les deux essais, à différentes P-values nominales.

I.3 Données complémentaires

I.3.1 Comparaison des modèles d'associations utilisés dans les deux populations étudiées

La **Figure 21** présente la distribution cumulée des p-values nominales observées pour les marqueurs détectés par les différents modèles (Naïf, Q, K, Q+K) jusqu'à un seuil de 0,05 dans chacune des deux populations. Comme dans le cas de l'échaudure des feuilles, le modèle naïf détecte un nombre très important d'associations pour les deux populations (142 et 230 respectivement en Guadeloupe et à la Réunion à un seuil $P < 0,01$).

Pour la population de la Réunion, les 3 modèles (Q, K et Q+K) montrent des courbes relativement bien superposées comme dans le cas de l'étude de la résistance à l'échaudure des feuilles présentée précédemment (voir graphiques associés au paragraphe I.3.1 aux pages 60). Dans le cas de la résistance au SCYLV on observe une réduction par un facteur 4 à 5 du nombre d'associations détectées par rapport au modèle naïf.

Pour la population de Guadeloupe, les modèles K et Q+K donnent un nombre comparable d'associations (trois fois plus faible qu'avec le modèle naïf à un seuil nominal de 0,01). Le modèle Q apporte une correction plus faible (2 fois moins d'associations qu'avec le modèle naïf à un seuil nominal de 0,01).

Les deux panels d'associations ont montré des comportements agronomiques légèrement différents pour la résistance au SCYLV, avec des fréquences d'individus extrêmes un peu supérieures à la Réunion (cf figures 1 et 2 pages 87 et 88). Outre des différences toujours possibles dans le déterminisme du caractère étudié (héritabilités éventuellement différentes), la correction proportionnellement plus faible apportée par le modèle Q dans le panel de Guadeloupe comparé à celui de la Réunion, peut en grande partie s'expliquer par i) des éventuelles différences de niveau de structuration entre les deux populations ii) la fraction de cette structure liée à la variation phénotypique prise en compte dans la matrice représentant le co-facteur Q (1 axe ACP expliquant 10 % de la variabilité phénotypique en Guadeloupe et 5 axes ACP expliquant 29 % de la variabilité à la Réunion).

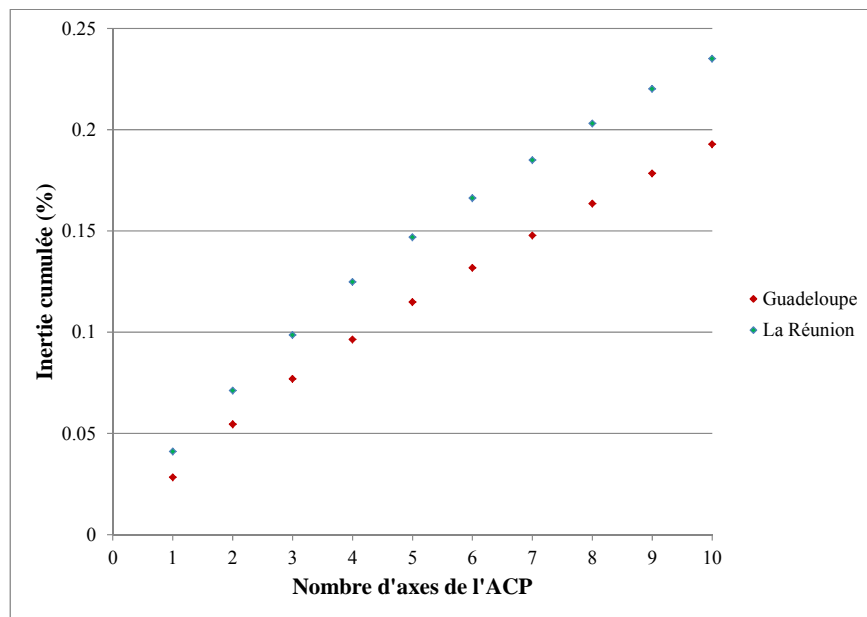


Figure 23: Comparaison de la structuration des deux populations (Guadeloupe et Réunion) - pourcentage d'inertie cumulée expliqué par les 10 premiers axes de l'ACP réalisée avec 467 marqueurs indépendants communs aux deux populations.

Afin d'analyser les éventuelles différences de structures entre populations sur une base comparable, une ACP a été réalisée, à partir d'un set de 467 marqueurs indépendants codés dans les deux populations (**Figure 23**). Les résultats montrent une structuration légèrement plus forte de la population de la Réunion par rapport à celle de Guadeloupe (10 axes de l'ACP expliquent 19,3 % et 23,5 % de la variation phénotypique, respectivement en Guadeloupe et à la Réunion).

Quand on examine les recouvrements de répertoires d'associations détectés entre modèles, on distingue un comportement différent entre les 2 panels, comme le montre la **Figure 22**. En Guadeloupe, le modèle Q détecte un nombre élevé d'associations spécifiques quelles que soient les p-values nominales examinées. En revanche les modèles K et Q+K ne détectent que très peu d'associations spécifiques à eux seuls ou exclusivement communes entre eux deux. Une forte proportion d'associations est détectée en commun par les trois modèles. A la Réunion, le nombre d'associations communes entre modèles observables sur une gamme de p-values nominales montre toujours la même tendance dans toutes les situations de p-values : les modèles Q et Q+K partagent toujours plus de marqueurs en commun que n'en partagent les modèles Q et K ou bien les modèles K et Q+K. Ces résultats indiquent que le co-facteur représentatif de la structure (Q) apporte une contribution importante dans le résultat du modèle global (K+Q) qui est théoriquement le plus performant.

I.3.2 Variation phénotypique expliquée par les marqueurs

La résistance au virus de la feuille jaune (SCYLV) observée dans des populations d'association de Guadeloupe (189 cultivars) et de la Réunion (184 cultivars) a été appréciée sur la base de l'incidence du virus (pourcentage de feuilles infectées). Nous avons effectué une analyse de variance des données de Guadeloupe, adaptée au dispositif expérimental utilisé (dispositif en 3 blocs complets randomisés), après une transformation normalisant les résidus ($\text{asin} \sqrt{\text{incidence}}$), en considérant un modèle mixte dans lequel l'effet clone était considéré comme aléatoire et l'effet bloc comme fixe (procédure Mixed de SAS (SAS Institute, 2008)).

L'héritabilité au sens large de l'incidence du SCYLV au niveau du dispositif, qui a été calculée à partir de l'estimation des variances clonales et résiduelles, s'élève à 0,74. Cette valeur relativement élevée indique que la caractérisation phénotypique de la population

d'association de Guadeloupe a été obtenue avec un degré de précision convenable.

Si on considère les haplotypes détectés à une GWER de 5% (*cf.* Table 2 de l'article, page 86) qui présentent moins de 5% de données manquantes, un ensemble de 16 haplotypes explique 55% de la variation de l'incidence du SCYLV en Guadeloupe (régression linéaire multiple ; ddl=130) ; et un ensemble de 12 marqueurs explique 48% de la variation de l'incidence du SCYLV à la Réunion (régression linéaire multiple, ddl=124).

Ces chiffres correspondent à un effet moyen de marqueurs de 3,4% et 4%, respectivement en Guadeloupe et à la Réunion. L'effet moyen des marqueurs observé en Guadeloupe concernant la résistance au SCYLV est légèrement supérieur à celui observé pour la résistance à l'échaudure des feuilles, qui varie entre 2,2% et 2,5% respectivement dans les essais 1 et 2 (*cf.* Table 3 de l'article correspondant page 57). Cette supériorité de l'effet moyen des marqueurs s'explique en partie par le fait que la population de Guadeloupe ségrège plus fortement pour le SCYLV que pour l'échaudure des feuilles, comme le montrent les distributions agronomiques respectives de la maladie bactérienne (*cf.* Figure 1, page 55) et de la maladie virale (*cf.* Figure 2, page 88) présentées dans les deux articles correspondants. Ces différences de distributions chez les cultivars modernes tiennent au fait que l'échaudure des feuilles est une maladie ancienne, intégrée depuis longtemps dans les critères d'amélioration et dont des symptômes sont facilement repérables lorsqu'ils s'expriment ; à la différence de la feuille jaune qui est une maladie émergente, encore très peu prise en compte dans les schémas de sélection, notamment en raison d'une symptomatologie délicate (jaunissements foliaires pouvant être confondus avec des causes de stress abiotiques et fréquentes situations de matériel sensible ne montrant pas de symptômes).

Dans l'ensemble, l'effet moyen des marqueurs liés à la résistance au SCYLV que nous obtenons dans nos populations d'associations (3,4% à 4%) se situe dans la moyenne de la gamme de variation trouvée dans divers études QTL relatives au rendement agronomique ou sucrier passées précédemment en revue (**Tableau 4**) : Ming *et al.*, 2001 (3,1%-4,7%) ; Ming *et al.*, 2002b (2,9%-9,3%) ; Hoarau *et al.*, 2002 (2,7%-5,6%) ; Aitken *et al.*, 2006 (1,6%-2%) ; Aitken *et al.*, 2008 (3,4% à 6,1%).

Cette similitude de l'ordre de grandeur de l'effet (R^2) moyen des marqueurs entre nos études d'associations (résistance au SCYLV) et les études QTL (divers caractères agronomiques) pourrait être a priori surprenante si l'on considère la différence de fonds

Tableau 9 : Résultats de régressions linéaires multiples réalisées sur les marqueurs significatifs ($P_{\text{GWER}} < 0.05$) indépendants, pour l'incidence du SCYLV à la Réunion et en Guadeloupe.

	Régression simple (marqueurs indépendants)			Régression pas à pas (sous-ensemble de marqueurs le plus parcimonieux)			
	Nombre de marqueurs	ddl	r^2	Nombre de marqueurs	ddl	r^2	Liste de marqueurs
Guadeloupe	16	130	0.55	10	141	0.51	D-1737 ; D-340 ; G-A-98 ; D-190 ; G-A-1093 ; G-A-232 ; D-817 ; G-A-1165 ; G-A-597 ; G-A-1728
Réunion	12	124	0,48	6	163	0,36	R-A-540 ; R-A-863 ; D-371 ; D-1342 ; D-1712 ; D-1220

généétique utilisés : près de 200 accessions dans nos études d'associations, contre uniquement une ou deux dans les études QTL (descendances bi- ou mono- parentales). Cette comparaison est intéressante en raison de l'existence de paramètres expérimentaux semblables (type de marqueurs identiques ou similaires en terme de polymorphisme ; tailles comparables des populations ayant servi à estimer les R^2) ; même s'il convient par ailleurs de la relativiser sur d'autres points (étendues des ségrégations agronomiques et densité du marquage pas nécessairement comparables).

Dans un contexte d'associations basées sur l'étude d'un large panel, le nombre d'allèles en ségrégation liés au déterminisme génétique du SCYLV est beaucoup plus élevé que celui pouvant exister dans toute étude QTL fondée sur un croisement biparental. En moyenne la taille individuelle des allèles utiles est probablement réduite du fait d'un effet de dilution de chaque allèle parmi un très grand nombre. Comme discuté dans notre premier article (*cf.* Rocher *et al.* page 52), la valeur appréciable des R^2 des marqueurs que nous trouvons ici associés à la résistance au SCYLV (3,4% à 4%) s'explique avant tout par la nature du marquage. Dans le cas d'une cartographie de QTL qui est par ailleurs toujours basée chez la canne à sucre sur les seuls marqueurs simplex, l'effet d'un marqueur correspond à l'effet moyen d'un seul allèle. En revanche, dans une étude d'association fondée sur un panel d'individus divers, l'effet d'un marqueur correspond à l'effet moyen d'un nombre inconnu de doses, plus ou moins important selon les loci présents et le niveau d'hétérozygotie.

Pour conclure, un calcul de régression multiple pas-à-pas réalisé avec l'ensemble des haplotypes révélés ($P_{\text{GWER}} < 0,05$) dans les panels de Guadeloupe et de La Réunion, nous a permis d'identifier des sous-ensembles de 10 et 6 haplotypes respectivement, qui permettent d'expliquer 51% (ddl=141) et 36% (ddl=163) de la variation de la résistance au SCYLV (**Tableau 9**). Cela correspond respectivement à une moyenne 5,1% et de 6% par haplotype.

Pour des perspectives d'applications en amélioration variétale, dans l'hypothèse d'allèles à effet additif, les haplotypes les plus intéressants parmi ceux révélés avec des effets comparables, sont ceux dont les effets visibles (R^2 individuel) correspondraient à un cumul de dosage le plus modeste possible, et ceux apparaissant en faible fréquence dans les populations étudiées.

Sur le premier critère le marquage dominant dont nous sommes tributaires ne permet pas de distinction entre haplotypes. En revanche sur le second critère, il est possible

Tableau 10 : Origine ancestrale des marqueurs associés à l'incidence du SCYLV à la Réunion et en Guadeloupe ($P_{GWER} < 0,05$) (Type D : marqueur DArT, A : marqueur AFLP ; Effet S: marqueur associé à la sensibilité, Type R: marqueur associé à la résistance).

Type	Réunion			Guadeloupe			Effet	Origine
	Haplotype	Marqueur	Fréquence	Haplotype	Marqueur	Fréquence		
D	R-1-19	D-212	0.11	G-1-11	D-212	0.09	R	<i>S. spontaneum</i>
D	R-1-19	D-52	0.18					-
D	R-1-19	D-817	0.11	G-1-11	D-817	0.09		<i>S. spontaneum</i>
A	R-1-19	R-A-903	0.3					-
D	R-1-13	D-1342	0.87				S	<i>S. spontaneum</i>
D	R-1-13	D-757	0.85					<i>S. spontaneum</i>
D				G-1-48	D-102	0.65	R	-
D				G-1-48	D-1737	0.62		-
D				G-1-34	D-1363	0.07	R	Non-spécifique
D				G-1-34	D-340	0.20		<i>S. spontaneum</i>
A				G-1-34	G-A-1776	0.14		-
D	R-1-36	D-371	0.12				R	<i>S. spontaneum</i>
D	R-1-79	D-1220	0.18				R	<i>S. spontaneum</i>
D	R-1-16	D-1275	0.58				S	-
D	R-1-16	D-1405	0.56					Non-spécifique
D	R-1-16	D-1433	0.57					-
D	R-1-16	D-1474	0.61					Non-spécifique
A	R-1-16	R-A-65	0.46					-
A	R-1-16	R-A-863	0.39					-
A	R-1-172	R-A-540	0.58				R	-
A	R-1-133	R-A-154	0.11				R	-
D	R-1-160	D-186	0.89				S	<i>S. spontaneum</i>
D	R-1-154	D-1820	0.07				R	-
D	R-1-182	D-1326	0.29				R	<i>S. spontaneum</i>
D	R-1-161	D-556	0.9				S	-
D	R-1-135	D-1712	0.94				S	<i>S. spontaneum</i>
A				G-1-147	G-A-1137	0.16	R	<i>S. spontaneum</i>
A				G-1-168	G-A-1266	0.32	R	<i>S. spontaneum</i>
A				G-1-166	G-A-401	0.12	R	-
A				G-1-167	G-A-597	0.20	R	-
A				G-1-156	G-A-1728	0.63	S	<i>S. spontaneum</i>
A				G-1-148	G-A-98	0.18	R	-
A				G-1-129	G-A-232	0.11	R	<i>S. spontaneum</i>
D				G-1-10	D-228	0.13	R	<i>S. spontaneum</i>
A				G-1-15	G-A-1093	0.15	R	<i>S. spontaneum</i>
A				G-1-15	G-A-693	0.29		-
A				G-1-20	G-A-1873	0.44	R	<i>S. spontaneum</i>
A				G-1-20	G-A-386	0.39		<i>S. spontaneum</i>
A				G-1-26	G-A-1758	0.23	S	-
A				G-1-26	G-A-472	0.35		<i>S. spontaneum</i>
D				G-1-37	D-310	0.14	R	<i>S. spontaneum</i>
A				G-1-37	G-A-513	0.11		<i>S. spontaneum</i>
D				G-1-37	D-993	0.52		<i>S. spontaneum</i>
A				G-1-40	G-A-1165	0.06	R	<i>S. spontaneum</i>
D				G-1-41	D-1720	0.61	S	Non-spécifique
D				G-1-52	D-1757	0.08	R	<i>S. spontaneum</i>
D				G-1-52	D-299	0.07		<i>S. spontaneum</i>
D				G-1-52	D-398	0.07		<i>S. spontaneum</i>
D				G-1-70	D-1056	0.14	R	<i>S. spontaneum</i>
D				G-1-70	D-1814	0.16		Non-spécifique
D				G-1-70	D-190	0.15		<i>S. spontaneum</i>

d'identifier 8 haplotypes de résistance de faible fréquence dans le panel de Guadeloupe [D-340 (0,20) ; G-A-98 (0,18) ; G-A-1093 (0,15) ; G-A-232(0,11) ; D-187 (0,09) ; G-A-1165 (0,06) ; G-A-597 (0,20) ; G-A-1728 (0,18)] et deux autres haplotypes résistants de faible fréquence dans le panel de La Réunion [D-371 (0,12) et D-1220 (0,18)] (**Tableau 10**).

I.3.3 Détermination de l'origine ancestrale spécifique des marqueurs

Suivant la méthodologie développée précédemment (*cf.* paragraphes I.3.2. pages 61 à 65), nous avons étudié l'origine ancestrale spécifique des haplotypes associés à la résistance au SCYLV. Sur les 30 haplotypes détectés à une P_{GWER} de 0,05 sur l'ensemble des deux panels d'association, 20 haplotypes sont d'origine *S. spontaneum*, 2 haplotypes ne sont spécifiques d'aucune des 2 espèces ancestrales et enfin 8 haplotypes sont d'origine indéterminée (**Tableau 10**). Ces résultats montrent que la variabilité de la résistance au SCYLV parmi les cultivars modernes est principalement attribuable à la ségrégation d'allèles présents dans la fraction de leur génome hérité de l'espèce sauvage (*S. spontaneum*).

Quatrième partie :

Conclusion, perspectives

I.1 Rappel du contexte de l'étude

Au cours de ce travail une analyse du déterminisme génétique de la résistance à deux maladies de la canne à sucre a été abordée à travers des études de génétique d'associations exploitant le déséquilibre de liaison appréciable existant parmi les cultivars modernes. Les travaux effectués reposent sur l'étude de deux larges panels de cultivars représentatifs de la diversité génétique actuellement utilisée dans les programmes d'amélioration.

Deux maladies à caractère systémique, d'impact économique important, ont été étudiées à travers des expérimentations réalisées au champ et conduites en Guadeloupe et à La Réunion. La résistance à l'échaudure des feuilles (*X. albilineans*) a été évaluée au champ (indice de sévérité) sur le seul panel de Guadeloupe, dans un contexte écologique favorisant l'expression de symptômes foliaires, consécutifs à la propagation de la bactérie par voie aérienne. La maladie virale de la feuille jaune (SCYLV), qui est transmise par la même espèce de puceron (*M. sacchari*) en Guadeloupe et à la Réunion, a été évaluée par une mesure de l'incidence du virus (test sérologique) sur les panels des deux îles qui comportaient 29 accessions communes.

Le travail réalisé résulte d'une collaboration entre les UMR AGAP (Guadeloupe et Montpellier) et PVBMT (Réunion). L'objectif principal consistait à tester la faisabilité d'approches de génétique d'association dans le contexte hautement polyploïde de la canne à sucre, moyennant le génotypage des deux panels de cultivars avec 3700 marqueurs polymorphes en moyenne, dont 1259 ont été codés en commun. Le but était de voir dans quelle mesure il était possible de marquer des facteurs de résistance à l'échaudure des feuilles et au SCYLV en mobilisant les méthodologies d'analyses fréquemment utilisées en génétique d'association chez les espèces moins complexes.

I.2 Résultats

I.2.1 Résultats généraux des modèles d'associations

- Le phénotypage des deux panels pour chaque maladie a été acquis dans des dispositifs expérimentaux et selon des méthodologies qui ont permis de bons niveaux de précision. Nous avons obtenu des ségrégations agronomiques appréciables, propices à des analyses génétiques. Les ségrégations plus larges observées pour le SCYLV comparativement à celles de l'échaudure des feuilles reflètent le fait que la première maladie est émergente, alors que la seconde est déjà prise en compte dans les programmes d'amélioration depuis longtemps ;
- Les modèles d'associations linéaires que nous avons utilisé, basés sur la prise en compte de cofacteurs d'analyses relatifs à la structure des panels (Q) et aux liens d'apparentements entre individus (K), permettent de diminuer plus ou moins fortement (selon les modèles) le nombre d'associations initialement détectées (par le modèle naïf), de façon comparable à ce qui est obtenu dans de nombreuses études d'associations sur des modèles de plantes plus simples (Zhao *et al.*, 2007 ; Ehrenreich *et al.*, 2009) ;
- Aux seuils des p-values nominales les plus intéressantes ($P < 0.005$), les recoupements des répertoires d'associations détectées entre les trois modèles (K, Q et Q+K), ou entre deux d'entre eux, varient fortement entre les deux panels étudiés pour le caractère de la résistance au SCYLV. Le panel de Guadeloupe montre une proportion plus élevée d'associations communes entre les trois modèles que celui de la Réunion. Ces résultats sont liés à un cumul de multiples causes possibles. Certaines sont identifiées (différents contextes épidémiologiques entre les deux îles, légères différences de degré de structuration entre panels ainsi que de la taille des ségrégations agronomiques, contribution non comparable de la structure des panels aux variations agronomiques) ; d'autres causes éventuelles restent inconnues.
- Un second filtre statistique (basé sur des tests de permutations) appliqué aux associations détectées dans un des modèles (Q) a permis d'identifier *in fine*, sur le panel de Guadeloupe, 23 associations significatives liées la résistance à l'échaudure des feuilles, à un risque de première espèce global sur le génome (P_{GWER}) de 5%, ainsi qu'un total de 30 associations significatives liés à la résistance au SCYLV sur l'ensemble des deux panels.

I.2.2 Variations agronomiques expliquées par le marquage et identification d'allèles d'intérêt prioritaires

- Avec un marquage de 3700 marqueurs polymorphes en moyenne par panel, les associations identifiées expliquent globalement entre 41% et 55% des variations agronomiques selon les caractères considérés ;
- Une étude comparant le profil moléculaire d'un petit groupe de clones *S. officinarum* avec un petit groupe de clones *S. spontaneum*, a permis de montrer qu'une grande majorité des marqueurs détectés en déséquilibre de liaison avec des allèles utiles dans nos panels de cultivars interspécifiques provenaient de l'espèce sauvage (*S. spontaneum*) ;
- En dépit du nombre élevé de fonds génétiques présents dans les panels étudiés (près de 200 cultivars), les associations trouvées expliquent un pourcentage appréciable de la variation phénotypique (R^2 de 2,2% à 2,5% pour la résistance à l'échaudure des feuilles et de 3,4% à 4% pour la résistance au virus). Ce pourcentage est comparable aux effets d'allèles simplex trouvées dans plusieurs études QTL antérieures, relatives à divers caractères agronomiques (rendement agronomique ou sucrier) et fondées sur des descendance mono- ou bi-parentales (Hoarau *et al.*, 2002 ; Reffay *et al.*, 2005 ; Aitken *et al.*, 2006 ; Aitken *et al.*, 2008 ; Nibouche *et al.*, 2011) ;
- Compte tenu de la constitution génétique des cultivars modernes marquée par la présence d'un faible nombre de chromosomes issus de *S. spontaneum* dans chacune des classes d'homéologie (entre 10% et 20% en moyenne), on peut penser que l'effet moyen de ces marqueurs provenant de l'espèce sauvage correspond à l'effet cumulé d'un nombre moyen de doses alléliques minimales (1 à 2 par cultivar possédant les marqueurs considérés) ;
- Compte tenu de la nature non locus-spécifique du marquage que nous avons employé, il est possible que certains marqueurs étiquettent plus d'un locus chez certains cultivars qui les possèdent. De nombreuses configurations génétiques sous-jacentes sont envisageables. Toutefois la valeur des R^2 de ces marqueurs hérités de l'espèce sauvage devrait correspondre à une estimation assez proche de la valeur réelle des allèles auxquels ils sont associés. Selon les configurations génétiques sous-jacentes, il est probable que les valeurs de R^2 correspondent soit à une légère surévaluation de l'effet de l'allèle dans certains cas, soit à une légère sous-évaluation dans d'autres. Par exemple dans

- l'hypothèse de l'existence de 2 copies d'un marqueur *S. spontaneum* chez les cultivars montrant un phénotype [présent], qui ne seraient liées ni en couplage (sur le même chromosome) ni en répulsion (sur 2 chromosomes homologues *S. spontaneum*), l'effet réel de l'allèle pourrait être légèrement sous-évalué si ce dernier ne se trouvait en déséquilibre de liaison qu'avec un seul des 2 marqueurs (existence de cultivars marqués pouvant ne pas posséder l'allèle). A l'inverse, une légère surestimation de l'effet du marqueur serait probable dans l'hypothèse d'un marquage spécifique d'un seul locus, si une majorité de cultivars de phénotype [présent] renfermaient deux chromosomes *S. spontaneum* homologues portant le même marqueur, chacun en déséquilibre de liaison avec un exemplaire de l'allèle.
- Parmi les marqueurs détectés, nous en avons identifié un certain nombre présents en fréquence faible à modeste dans chacun des deux panels, qui sont porteurs d'effets de résistance (8 marqueurs pour l'échaudure dans le panel de Guadeloupe ; 8 et 2 marqueurs pour la résistance au SCYLV dans les panels de Guadeloupe et de la Réunion respectivement). Compte tenu de leur fréquence dans les panels étudiés, il est très probable que certains de ces marqueurs étiquettent des QTL à effets forts, voir majeurs.

I.2.3 Comparaison des associations détectées dans les deux panels

Cet exercice de comparaison concerne uniquement la résistance au SCYLV. Sa portée est limitée par le nombre modeste de marqueurs codés en communs entre les deux panels (1259). Les résultats obtenus font apparaître une interaction cultivar x site significative pour les 29 cultivars communs aux deux panels. Cette interaction s'explique principalement par i) des différences de diversité génétique du virus entre les 2 îles (3 génotypes viraux présents en Guadeloupe, et 2 à La Réunion) ii) des prévalences de génotypes viraux non comparables iii) des différences concernant la capacité des populations de pucerons à transmettre le virus, liée à des facteurs écologiques et/ou génétiques (Gray and Gildow, 2003) .

En raison des différences épidémiologiques notables entre les deux sites, nous n'avons détecté que cinq associations communes entre les deux panels, dont quatre correspondent à des allèles de résistance, vraisemblablement des facteurs de résistance généraux et peu spécifiques d'un génotype viral particulier.

I.3 Perspectives

I.3.1 Un retour à l'étude de ressources génétiques élit

La découverte d'allèles de résistance d'origine *S. spontaneum* en fréquence faible dans les panels de cultivars, indique que certains cultivars seraient porteurs d'allèles de résistance importants, voir majeurs. Pour le vérifier il faudrait entreprendre des études QTL. Cela nécessiterait d'abord de réaliser un certain nombre de croisements bi-parentaux ciblés entre des cultivars résistants et des cultivars sensibles ; d'expérimenter sur le terrain les descendances candidates ainsi produites et de toutes les phénotyper en conditions naturelles pour chacune des deux maladies. Ce travail d'études QTL représente un effort conséquent qui implique la mobilisation de ressources importantes. Toutefois l'identification des cultivars élites pouvant contenir des gènes majeurs constitue une étape indispensable pour espérer construire une amélioration génétique raisonnée de la résistance aux maladies en général, et au SCYLV et à l'échaudure des feuilles en particulier.

L'échaudure des feuilles est une maladie prise en compte depuis longtemps dans les critères de sélection, dans la plupart des pays qui présentent des écologies favorisant son développement. La sélection effectuée repose uniquement sur une base phénotypique. Les études génétiques sont rares (Bressiani *et al.*, 2007). La sélection phénotypique contre l'échaudure des feuilles ne pourrait être efficace que dans les seules écologies de culture qui garantiraient en permanence l'expression de symptômes sur le matériel sensible soumis en routine à la sélection. Ces écologies de culture, si elles existent, doivent être rares, car les régions tropicales et intertropicales où sont localisées les programmes de sélection sont souvent caractérisées par de fortes variations climatiques interannuelles. Il en résulte que dans de nombreux pays, il est fréquent de découvrir tardivement dans le processus de sélection des clones sensibles devant être rejetés malgré leurs bonnes performances agronomiques (JY Hoarau pers. com.). Ces éliminations tardives qui limitent l'efficacité des schémas de sélection pourraient être atténuées si les travaux d'amélioration reposaient sur une meilleure connaissance des bases génétiques contrôlant la résistance à l'échaudure des feuilles.

Le contexte d'amélioration génétique de la résistance au SCYLV est encore davantage critique. Il s'agit d'une maladie émergente, pour laquelle il n'existe encore aucun

programme d'amélioration génétique dans aucun pays (JY Hoarau pers.com.). La sélection phénotypique est totalement inefficace, compte tenu de la possibilité de confusion des symptômes de jaunissement foliaires avec l'effet de stress biotiques (Rassaby *et al.*, 2003), et les fréquentes situations asymptomatiques (clones sensibles passant inaperçus) (Comstock *et al.*, 1998 ; Izaguirre-Mayoral *et al.*, 2002). Il en résulte, dans de nombreux pays où le SCYLV est présent, une proportion élevée de clones sensibles dans les stades avancés de sélection faisant l'objet de tests sérologiques. De surcroît une étude récente sur la diversité génétique mondiale du puceron (*M. sacchari*) menée par l'UMR PVBMT a montré l'existence d'un génotype de puceron très répandu dans le monde qui traduirait un potentiel adaptatif élevé (Fartek *et al.*, b, soumis).

Le Cirad a récemment produit quelques dizaines de descendance en Guadeloupe à partir de combinaisons bi-parentales entre des accessions appartenant au panel d'associations local et ayant pu fleurir de façon synchrone (D. Roques com. pers.). Ces populations expérimentales pourront être utiles à des études QTL sur plusieurs caractères, dont la résistance au SCYLV et à l'échaudure des feuilles, si les conditions d'infections naturelles sont favorables.

I.3.2 A la recherche de nouvelles techniques de marquage

Un des enjeux des analyses génétiques à venir réside dans l'accès à des techniques de marquage innovantes permettant un marquage plus dense du génome de la canne à sucre. L'augmentation de la couverture en marqueurs est rendue difficile chez la canne à sucre à cause de l'étendue du génome (10 Gb) et de sa redondance élevée due à son organisation polyploïde (12 copies homéologues). La technologie DArT a déjà apporté un progrès notable en permettant l'acquisition rapide de marqueurs polymorphes et en démontrant l'intérêt d'une stratégie GWA (Genome Wide Association) pour identifier les régions génomiques impliquées dans des caractères d'intérêt agronomique. Toutefois cette technique ne produit qu'environ 500 marqueurs cartographiables sur une carte et environ 1500 marqueurs polymorphes sur une population de près de 200 cultivars. Schématiquement, nous aurions besoin de multiplier au moins par 10 ce nombre de marqueurs pour optimiser la couverture du génome (notamment les zones héritées de *S. officinarum*) et l'efficacité des études de génétique d'association ou de sélection assistée par marqueur, et dans le futur, des approches de sélection génomique visant à prédire la valeur phénotypique des individus soumis à sélection.

De nouvelles technologies permettant d'obtenir une couverture du génome plus importante sont en train d'émerger. Les méthodes de génotypage basées sur la détection de SNPs, méthodes extrêmement efficaces pour les études d'association chez des espèces de plantes plus simples comme le riz (Zhao *et al.*, 2011), commencent à être développées chez la canne à sucre (Bundock *et al.*, 2009 ; Marconi *et al.*, 2011). Mais ces études restent coûteuses, notamment à cause de la forte proportion de SNPs multidoses révélés (Grivet *et al.*, 2003 ; données non publiées), alors que seuls les marqueurs simple-dose sont exploitables en cartographie génétique chez la canne. Il s'avère donc nécessaire de développer des technologies de marquage haut débit peu coûteuses pour révéler un maximum d'allèles SNP simplex. Récemment, grâce au développement des NGS (New Generation Sequencing), le coût du séquençage a considérablement diminué et devrait diminuer encore, permettant ainsi un changement d'échelle dans le marquage moléculaire. Le génotypage par séquençage (GBS, Genotyping By Sequencing) devient envisageable chez la canne à sucre. Une autre avancée méthodologique consistant à réduire la proportion de génome séquencé permet d'espérer pouvoir appliquer le GBS chez les espèces polyploïdes à gros génome et d'améliorer la proportion de SNPs simple dose révélés. Par exemple, la technologie RAD (Restriction-site Associated DNA) proposée par Miller *et al.* (2007) consiste à n'amplifier que les séquences adjacentes à un site de restriction enzymatique. Cette réduction permet d'amplifier un plus grand nombre de séquences pour un site donné, donc d'augmenter la couverture de séquençage pour chaque site, ainsi que la probabilité de détecter un allèle simplex. La probabilité que les mêmes sites de différents échantillons d'ADN soient séquencés est elle aussi augmentée. La ligation d'adaptateurs spécifiques sur les fragments d'ADN digérés permet un multiplexage des ADN avant séquençage NGS. La technologie RAD a été appliquée à des études de cartographie génétique et de détection de QTL chez un large panel d'espèces, dont des Poacées et des espèces pour lesquelles aucun génome de référence n'était disponible (Chutimanitsakun *et al.*, 2011 ; Pfender *et al.*, 2011 ; Rowe *et al.*, 2011). Deux équipes du Cirad (Montpellier et La Réunion) sont en train de mettre en place une collaboration avec une équipe américaine spécialiste d'une technologie RAD, modifiée et approuvée chez d'autres espèces de plantes comme le maïs (Elshire *et al.*, 2011), afin de tester et d'optimiser le GBS chez la canne et d'identifier plus précisément de nouveaux SNPs potentiellement liés à des facteurs de résistance pour affiner leur cartographie génétique.

Références bibliographiques

- Abdurakhmonov I.Y., Saha S., Jenkins J.N., Buriev Z.T., Shermatov S.E., Scheffler B.E. Pepper A.E. Yu J.Z., Kohel R.J. Abdugarimov A. (2009)** Linkage disequilibrium based association mng of fiber quality traits in *G. hirsutum* L. variety germplasm. *Genetica* **136**:401-417.
- Abu Ahmad Y., Rassaby L., Royer M., Borg Z., Braithwaite K., Mirkov T., Irej M., Perrier X., Smith G. et Rott P. (2006a)** Yellow leaf of sugarcane is caused by at least three different genotypes of *Sugarcane Yellow Leaf Virus*, one of which predominates on the island of Réunion. *Archives of virology* **151**:1355-1371.
- Abu Ahmad Y., Royer M., Daugrois J.H., Costet L., Lett J.M. Victoria J.I. Girard J.C., Rott P. (2006b)** Geographical distribution of four *Sugarcane Yellow Leaf Virus* genotypes. *Plant Disease* **90**:1156-1160.
- Abu Ahmad Y., Costet L., Daugrois J.H., Nibouche S., Letourmy P., Girard J.C. et Rott P. (2007)** Variation in infection capacity and in virulence exists between genotypes of *Sugarcane Yellow Leaf Virus*. *Plant disease* **91**(3):253-259.
- Achleitner A., Tinker N., Zechner E., Buerstmayr H. (2008)** Genetic diversity among oat varieties of worldwide origin and association of AFLP markers with quantitative traits. *Theoretical and applied genetics* **117**:1041-1053.
- Agrama H.A., Eizenga G.C., Yan W. (2007)** Association mapping of yield and its components in rice. *Molecular breeding* **19**:341-356.
- Aitken K., Jackson P., McIntyre C. (2005)** A combination of AFLP and SSR markers provides extensive map coverage and identification of homo(eo)logous linkage groups in a sugarcane cultivar. *Theoretical and Applied Genetics* **110**: 789-801.
- Aitken K., Jackson P., McIntyre C. (2006)** Quantitative trait loci indentified for sugar related traits in sugarcane (*Saccharum* spp.) cultivar x *Saccharum officinarum* population. *Theoretical and Applied Genetics* **112**: 1306-1317.
- Aitken K.S., Hermann S., Karno K., Bonnett G.D., McIntyre L.C., Jackson P.A. (2008)** Genetic control of yield related stalk traits in sugarcane. *Theoretical and Applied Genetics* **117**:1191-1203
- Akbar W., Showler A.T., Reagan T.E., White W.H. (2010)** Categorizing Sugarcane resistance to the Sugarcane Aphid and Yellow Sugarcane Aphid (*Hemiptera: Aphidiae*). *Journal of economical society of America* **103**(4):1431-1437.
- Akbar W., Showler A.T., Beuzelin J.M., Reagan T.E., Gravois K.A. (2011)** Evaluation of aphid resistance among sugarcane cultivars in Louisiana. *Journal of economical society of America* **104**(4):699-704.
- Aljanabi S.M., Honeycutt R.J., McClelland M., Sobral B.W.S. (1993)** A genetic-linkage map of *Saccharum spontaneum* L. 'SES 208'. *Genetics* **134**: 1249-1260.
- Aljanabi S.M., Parmessur Y., Kross H., Dhayan S., Saumtally S., Ramdoyal K., Autrey L.J.C., Dookun-Saumtally A. (2007)** Identification of a major quantitative trait locus (QTL) for yellow spot (*Mycovellosiella koepkei*) disease resistance in sugarcane. *Molecular Breeding* **19**: 1-14.
- Alwala S., Kimbeng C., Veremis J., Gravois K. (2008)** Linkage mapping and genome analysis in a *Saccharum* interspecific cross using AFLP, SRP and TRAP markers. *Euphytica* **164**:37-51.

- Andru S., Pan Y-B., Thongthawee S., Burner D., Kimbeng C. (2011)** Genetic analysis of the sugarcane (*Saccharum* spp.) cultivar ‘LCP 85-384’. I. Linkage mapping using AFLP, SSR, and TRAP markers. *Theoretical and applied genetics* **123**: 77-93.
- Aranzana M.J., Kim S., Zhao K., Bakker E., Horton M., Jakob K., Lister C., Molitor J., Shindo C., Tang C., Toomajian C., Traw B., Zheng H., Bergelson J., Dean C., Marjoram P., Nordborg M. (2005)** Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. *Plos Genetics* **1**(5):531-539.
- Arceneaux G. (1965)** Cultivated sugarcanes of the world and their botanical derivation. *Proc. Int. Soc. of Sugarcane Technol.* **12**: 844-854.
- Artschwager E. and Brandes W. (1958)** Sugarcane (*Saccharum officinarum* L.): origin, classification, characteristics and description of representative clones. *Washington DC, United States department of agriculture*.
- Asnaghi C., Roques D., Ruffel S., Kaye C., Hoarau J. Y., Telismart H., Girard J.C., Raboin L.M., Risterucci A.M., Grivet L., D’Hont A. (2004)** Targeted mapping of a sugarcane rust resistance gene. *Theoretical and applied genetics* **108**: 759-764.
- Autrey L.J.C., Saumtally S., Dookun A., Sullivan S., Dhayan S. (1995)** Aerial transmission of the leaf scald pathogen, *Xanthomonas albilineans*. *Proceedings of the International Society of Sugar Cane Technologists* **21**:508-526.
- Bechet G.R., Bailey R.A., Nuss K.J. (1992)** Screening of new sugarcane clones for resistance to leaf scald (*Xanthomonas albilineans*). *Proc. South African Sugar technol. Association* **66**:82-86.
- Berding N. and Hurney P. (2005)** Flowering and lodging, physiological-based traits affecting cane and sugar yield – What do we know of their control mechanisms and how do we manage them? *Field Crops Research* **92**: 261-275.
- Berndes G. (2002)** Bioenergy and water – the implications of large-scale bioenergy production for water use and supply. *Global environmental change* **12**: 253-271
- Bhat S.R. and Gill B.S. (1985)** The implication of 2n egg gametes in nobilisation and breeding of sugarcane. *Euphytica* **34**:377-384.
- Birch R.G. (2001)** *Xanthomonas albilineans* and the antipathogenesis approach to disease control. *Molecular Plant Pathology* **2**(1):1-11.
- Blume H. (1985)** Geography of sugarcane. Environmental, structural and economical aspects of cane sugar production. *Verlag -Bartens editions, Berlin*, 371 pages.
- Brault V., Herrbach E., Hauser S., Lemaire O. (2001)** Les *luteoviridae* : propriétés biologiques et évolution. *Virologie* **5**(1) :9-21.
- Brault V., Bergdoll M., Mutterer J., Prasad V., Pfeffer S., Erdinger M., Richards K.E. Ziegel-Graff V. (2003)** Effect of point mutations in the major capsid protein of beet western yellows virus on capsid formation, virus accumulation and aphid transmission. *Journal of virology* **77**:3247-3256.

- Brault V., Pérignon S., Reinbold C., Erdinger M., Scheidecker D., Herrbach E., Richards K., Ziegler-Graff V. (2005)** The *Polerovirus* minor capsid protein determines vector specificity and intestinal tropism in the aphid. *Journal of virology* **79**:9685-9693.
- Bremer G. (1961)** Problems in breeding and cytology of sugarcane. I. A short history of sugarcane breeding - the original forms of *Saccharum*. *Euphytica* **10**: 59-78.
- Bressiani J.A., Sanguino A., Burnquist W.L., Vencovsky R., Da Silva G. (2007)** Breeding sugarcane for leaf scald resistance: a genetic study. *Journal American Society Sugar Cane Technol.* **27**:15-22.
- Bundock PC, Elliott FG, Ablett G, Benson AD, Casu RE, Aitken KS, Henry RJ. (2009)** Targeted single nucleotide polymorphism (SNP) discovery in a highly polyploid plant species using 454 sequencing. *Plant Biotechnol J* **7**: 347-354.
- Burner D.M. and Legendre B.L. (1993)** Chromosome transmission and meiotic stability of sugarcane (*Saccharum* spp.) hybrids derivatives. *Crop Sci.* **33**:600-606.
- Butterfield M., D'Hont A., Berding N. (2001)** The sugarcane genome: a synthesis of current understanding, and lessons for breeding and biotechnology. *Proc. South Afr. Technol. Ass.* **75**:1-5.
- Butterfield M.K. (2007)** Marker assisted breeding in sugarcane: a complex polyploidy. *University of Stellenbosch , South Africa– PhD thesis.*
- Büttner D. and Bonas U. (2010)** Regulation and secretion of *Xanthomonas* virulence factors. *FEMS Microbiol.Rev.* **34**:107-133.
- Cai Q. Aitken K., Deng H., Chen X.W., Fu C., Jackson P.A., McIntyre C.L. (2005)** Verification of the introgression of *Erianthus arundinaceus* germplasm into sugarcane using molecular markers. *Plant Breeding* **124**:322-328.
- Chaillet H. – in : Rott P., Bailey R.A., Comstock J.C., Croft B.J. et Saumtally A.S. (2000)** A guide to sugarcane diseases. *La librairie du CIRAD, Montpellier, France.* p14.
- Champoiseau P., Daugrois J.H., Girard J.-C., Royer M., Rott P. (2006a)** Variation in albicidin biosynthesis genes and in pathogenicity of *Xanthomonas albilineans*, the sugarcane leaf scald pathogen. *Phytopathology* **96**:33-45.
- Champoiseau P., Daugrois J.-H., Pierretti I., Cociancich S., Royer M., Rott P. (2006b)** High variation in pathogenicity of genetically closely related strains of *Xanthomonas albilineans*, the sugarcane leaf scald pathogen in Guadeloupe. *Phytopathology* **96**:1081-1091.
- Champoiseau P., Rott P., Daugrois J.H. (2009)** Epiphytic populations of *Xanthomonas albilineans* and subsequent sugarcane stalk infection are linked to rainfall in Guadeloupe. *Plant Disease* **93**:339-346.
- Chen C. T., Lin C. P. et Liang Y. G. (1993).** Leaf scald of sugarcane in Taiwan. *Taiwan Sugar* **40**:8-16.

- Ching A., Caldwell K.S., Jung M., Dolan M., Smith P.S., Tingey S., Morgante M., Rafalski A.J. (2002)** SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genetics* **3**:19.
- Churchill G. and Doerge R. (1994)** Empirical threshold values for quantitative trait mapping. *Genetics* **138**:963-971.
- Chutimanitsakun Y, Nipper RW, Cuesta-Marcos A, Cistue L, Corey A, Filichkina T, Johnson EA, Hayes PM. (2011)** Construction and application for QTL analysis of a Restriction Site Associated DNA (RAD) linkage map in barley. *BMC Genomics* **12**:4.
- Comstock J.C. and Shine J.M. (1992)** Outbreak of Leaf Scald of Sugarcane, caused by *Xanthomonas albilineans*, in Florida. *Plant disease* **76**:426.
- Comstock J.C. and Irey M.S. (1992)** Detection of the sugarcane leaf scald pathogen, *Xanthomonas albilineans*, using tissue blot immunoassay, ELISA, and isolation techniques. *Plant Disease* **76**:1033-1035.
- Comstock J.C., Irey M.S., Lockhart B.E.L., Wang Z.K. (1998)** Incidence of yellow leaf syndrome in CP cultivars based on polymerase chain reaction and serological techniques. *Sugar Cane* **4**:21-24.
- Comstock J.C. (2001)** Foliar Symptoms of Sugarcane Leaf Scald. *Sugar journal*. September:23-31.
- Comstock J.C., Miller J.D., Schnell R.J. (2001)** Incidence of *Sugarcane Yellow Leaf Virus* in clones maintained in the world collection of sugarcane and related grasses at the US National Repository in Miami. *Florida. Sugar Tech* **3**(4):128-133.
- Crossa J., Burgueño J., Dreisigacker S., Vargas M., Herrera-Foessel S., Lillemo M., Singh R., Trethowan R., Warburton M., Franco J., Reynolds M., Crouch H., Ortiz R. (2007)** Association analysis of historical Bread Wheat germplasm using additive genetic covariance of relatives and population structure. *Genetics* **177**:1889-1913.
- Costet L., Raboin L.M., Payet M., D'Hont A., Nibouche S. (submitted)** Mapping of a major QTA for resistance to the yellow leaf disease of sugarcane caused by the *Sugarcane Yellow Leaf Virus* (*Luteoviridae* family).
- D'Hont A., Lu Y.H., Leon D.G., Grivet L., Feldmann J.C., Lanaud C., Glaszmann J.C. (1994)** A molecular approach to unraveling the genetics of sugarcane, a complex polyploidy of the andropogoneae tribe. *Genome* **37**:222-230.
- D'Hont A., Grivet L., Feldmann P., Rao S., Berding N., Glaszmann J.C. (1996)** Characterization of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. *Mol. Gen. Genet.* **250**: 405-413.
- D'Hont A., Ison D., Alix K., Roux C. and Glaszmann J.C. (1998)** Determination of basic chromosome numbers in the genus *Saccharum* by physical mapping of ribosomal RNA genes. *Genome* **41** : 221-225.
- D'Hont A, Paulet F, Glaszmann JC (2002)** Oligoclonal interspecific origin of 'North Indian' and 'Chinese' sugarcane. *Chromosome Res* **10**:253–262

- D'Hont A. (2005)** Unravelling the genome structure of polyploids using GISH and FISH; examples of sugarcane and banana. *Cytogenetic and genome research* **109**:27-33.
- D'Hont A., Mendes Souza G., Menossi M., Vincentz M., Van-Sluys M.A., Glaszmann J.C., Ulian E. (2008)** Sugarcane: a major source of sweetness, alcohol and bio-energy. In: Moore P., Ming R., *Genomic of Tropical Crop Plants* – Springer- 483-514.
- D'Hoop B.B., Paulo M.J., Ma,k R.A., van Eck H.J., van Eeuwijk F.A. (2008)** Association mapping of quality traits in potato (*Solanum tuberosum* L.). *Euphytica* **161**:47-60.
- Da Silva J.A., Sorrells M.E., Burnquist W.L., Tanksley S.D. (1993)** RFLP linkage map and genome analysis of *Saccharum spontaneum*. *Genome* **36**: 782-791.
- Da Silva J., Honeycutt R .J., Burnquist W., Al-Janabi S.M., Sorrells M.E., Tanksley S.D., Sobral B.W.S. (1995)** *Saccharum spontaneum* L. 'SES 208' genetic linkage map combining RFLP and PCR-based markers. *Molecular Breeding* **1**: 165-179.
- Daniels J., Smith P., Paton N., Roach B. (1975)** The origin of the genus *Saccharum*. *Sugarcane Breeding Newsletter* **36**:24-39.
- Daniels J. and Roach B. (1987)** Taxonomy and evolution. In: Heinz D.J. ed., *Sugarcane Improvement trough breeding*. New-York: Elsevier Press, 7-84.
- Daugrois J.H., Saed Mohamed B., Rott P. (1995)** Screening sugarcane germplasm for resistance to leaf scald disease by analysis of pathogen population densities. *Phytopathology* **85**(12):1562.
- Daugrois J.H., Grivet L., Roques D., Hoarau J.Y.,Lombard H., Glaszmann J.C., D'Hont A. (1996)** A putative major gene for rust resistance linked with a RFLP marker in sugarcane cultivar R570. *Theoretical and applied Genetics* **92**:1059-1064.
- Daugrois J.H., Costet L., Rott P. (1998)** Resistance to leaf scald disease in wild relatives of sugarcane analyzed by pathogen population densities. *7th International Congres of Plant Pathology – Edinburgh, Scotland – 9th-16th august*.
- Daugrois J.H, Champoiseau P., Costet L., Boisne-Noc R., Rott P. (2003)** Aerial contamination of sugarcane in Guadeloupe by two strains of *Xanthomonas albilineans*. *European Journal of plant pathology* **109**: 445-458.
- Daugrois J.H., Edon-Jock C., Bonoto S., Vaillant J., Rott P. (2011)** Spread of *Sugarcane Yellow Leaf Virus* in initially disease-free sugarcane is linked to rainfall and host resistance in the humid tropical environment of Guadeloupe. *Eur. J. Plant Pathol.* **129**:71-80.
- Daugrois JH, Boisne-Noc R, Champoiseau P, Rott P (in press).** The revisited infection cycle of *Xanthomonas albilineans*, the causal agent of leaf scald of sugarcane. *Functional Plant Science and Biotechnology - Special issue: Sugarcane Pathology*.
- Davis M. J., Rott P., Warmuth C. J., Chatenet M., and Baudin P. (1997)** Intraspecific genomic variation within *Xanthomonas albilineans*, the sugarcane leaf scald pathogen. *Phytopathology* **87**:316-324.

- Diaz M., Peralta E.L., Iglesias A., Pazos V., Carvajal O., Perez J., Gigliotti E.A., Gagliardi P.R., Wendland A. et Camargo L.E.A. (2001).** *Xanthomonas albilineans* haplotype B responsible for a recent sugarcane leaf scald disease outbreak in Cuba. *Plant Disease* **85**:334.
- Doerge R. and Churchill G. (1996)** Permutation tests for multiple loci affecting a quantitative character. *Genetics* **142**:285-294.
- Dogimont C., Bendamane A., Chovelon V., Boissot N. (2010)** Host plant resistance to aphids in cultivated crops: Genetic and molecular bases, and interactions with aphid populations. *C.R. Biologies* **333**:566-573.
- Edon C. (2008)** Le virus de la Feuille Jaune de la Canne à Sucre : Spécificités de la dissémination en Guadeloupe et en Martinique. *Thèse de doctorat de l'université des Antilles et de la Guyane*.
- Edon-Jock C., Vaillant P., Fernandez J., Girard J.C., Daugrois J.H. (2007)** Status of *Sugarcane Yellow Leaf Virus* in commercial fields and risk assessment in Guadeloupe. *Proceedings International Society of sugar Cane Technologists*. **26**:995-1004.
- Egan B.T. (1972)** Breeding for resistance to leaf scald disease. *Proc. Int. Soc. of Sugarcane technology* **14**: 920-924.
- Ehrenreich I.M., Hanzawa Y., Chou L., Roe J.L., Kover P.X., Purugganan M.D. (2009)** Candidate gene association mapping of *Arabidopsis* Flowering time. *Genetics* **183**:325-335.
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE.(2011)** A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **6**: e19379.
- Ersoz E.S., YU J. et Buckler E.S. (2007)** Application of linkage disequilibrium and association mapping in crop plants. In : *Genomic assisted crop improvement*, Varshney R.K and Tuberosa R. (editeurs), Springer, **1**: 97-119.
- Fartek B., Nibouche S., Turpin P., Costet L., Reynaud B. (a)** Mechanism of resistance to *Melanaphis sacchari* in the sugarcane cultivar R536. (Soumis à la revue *Entomologia experimentalis et applicata*).
- Fartek B., Mississippi S., Nibouche S., Costet L., Delatte H., Reynaud B. (b).** Molecular diversity of the sorghum and sugarcane aphids from the *Melanaphis sacchari*. (Soumis).
- Fauconnier R. (1991)** La canne à sucre. *Collection Le technicien d'agriculture tropicale. Editions Maisonneuve et Larose*. Paris, 163 pages.
- Flint-Garcia S.A., Thornsberry J.M. et Buckler E.S. (2003)** Structure of linkage disequilibrium in plants. *Annual Review of Plant Biology* **54**:357–74.
- Flint-Garcia S.A., Thuillet A.C., Yu J., Pressoir G., Romero S.M., Sharon E.M., Doebley J., Kresovich S., Goodman M.M., Buckler E.S. (2005)** Maize association population: a high-resolution platform for quantitative trait locus dissection. *Plant J* **44**:1054–1064.

- Fontaniella B., Vincente C., Legaz E.M., de Armas R., Rodriguez C.W., Martinez M., Pinon D., Acevedo R., Solas M.T. (2003)** Yellow leaf syndrome modifies the composition of sugarcane juices in polysaccharides, phenols and polyamines. *Plant Physiol. Biochem.* **41**:107-1036.
- Garcia A., Kido E., Meza A., Souza H., Pinto L., Pastina M., Leite C., da Silva J., Ulian E., Figueira A., Souza A. (2005)** Development of an integrated genetic map of a sugarcane (*Saccharum* spp.) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases. *Theoretical and applied genetics* **112**:298-314.
- Gonçalves M., Vega J., Oliveira J.G., Gomes M.M.A. (2005)** Sugarcane Yellow Leaf Virus infection leads to alterations in photosynthetic efficiency and carbohydrate accumulation in sugarcane leaves. *Fitopatol. Bras.* **30**:10-16.
- Gray S. and Gildow F. (2003)** Luteovirus-aphid interactions. *Annual review of phytopathology* **41**:539-566.
- Grivet L. (1995)** Marquage moléculaire chez la canne à sucre (*Saccharum* spp.) : Décomposition d'une structure génétique complexe et application à l'amélioration variétale. *Université de Paris XI, France – Thèse de doctorat.*
- Grivet L., D'Hont A., Roques D., Feldmann P., Lanaud C., Glaszmann J.C. (1996)** RFLP mapping in cultivated sugarcane (*Saccharum* spp.): Genome organization in a highly polyploid and aneuploid interspecific hybrid. *Genetics* **142**: 987-1000.
- Grivet L. and Arruda P. (2001)** Sugarcane genomics: depicting the complex genome of an important tropical crop. *Current opinion in plant biology* **5**:122-127.
- Grivet L., Glaszmann J.C., Vincentz M., da Silva F., Arruda P. (2003)** ESTs as a source for sequence polymorphism discovery in sugarcane: example of the Adh genes. *Theor Appl Genet* **106**:190-197.
- Grivet L., Daniels C., Glaszmann J.C., D'Hont A. (2004)** A review of recent molecular genetics evidence for sugarcane evolution and domestication. *Ethnobotany Research and Applications* **2**:9-17.
- Grivet L., Glaszmann J.C., D'Hont A. (2006)** Molecular evidence of sugarcane evolution and domestication. In: Motley Timothy J. , Zerega Nyree C. Hugh (eds). Darwin's harvest :New approaches to the origins, evolution, and conservation of crops. *New York Columbia University Press* 49-66.
- Guimaraes C.T., Honeycutt R.J., Sills G.R., Sobral B.W.S. (1999)** Genetic maps of *Saccharum officinarum* L. and *Saccharum robustum*. Brandes & Jew. Ex Grassl. *Genetics and Molecular Biology* **22**: 125-132.
- Habib I. (2008)** Le colonialisme – dans: Histoire de l'humanité, Volume V : 1492-1789. Editions UNESCO. ISBN 978-92-3-202814-3. pp143-145.

- Hoarau J.Y., Offmann B., D'Hont A., Risterucci A.M., Roques D., Glaszmann J.C., Grivet L. (2001)** Genetic dissection of a modern cultivar (*Saccharum* spp). I. Genome mapping with AFLP. *Theoretical and Applied Genetics* **103**:84–97.
- Hoarau J.Y., Grivet L., Offmann B., Raboin L.M., Diorflar J.P., Payet J., Hellmann M., D'Hont A., Glaszmann J.C. (2002)** Genetic dissection of a modern sugarcane cultivar (*Saccharum* spp). II. Detection of QTLs for yield components. *Theoretical and Applied Genetics* **105**:1027–1037.
- Holder D.G., Irey M.S., Davidson R.W., Hu C.J. (2000)** Yield losses caused by leaf scald in a commercial field of sugarcane cultivar CP80-1743. *Journal American Society of Sugar Cane Technologists* **20**:88-93.
- Hoy J.W. (1994)** Sugarcane leaf scald distribution, symptomatology and effects on yield in Louisiana. *Plant Disease* **78**:1083-1087.
- Isakeit T. and Irvine J. E. (1995)** First report of leaf scald, caused by *Xanthomonas albilineans*, of sugarcane in Texas. *Plant Disease* **79**:860.
- Izaguirre-Mayoral M.L., Carballo O., Alceste C., Romano M., Nass H.A. (2002)** Physiological performance of asymptomatic and yellow leaf syndrome-affected sugarcanes in Venezuela. *J. Phytopathology* **150**:13-19.
- Jannoo N., Grivet L., Dookun A., D'Hont A., Glaszmann J.C. (1999a)** Linkage disequilibrium among modern sugarcane cultivars. *Theoretical and Applied Genetics* **99**:1053-1060.
- Jannoo N., Grivet L., Seguin M., Paulet F., Domaingue R., Rao P.S., Dookun A., D'Hont A., Glaszmann J.C. (1999b)** Molecular investigation of the genetic base of sugarcane cultivars. *Theoretical and Applied Genetics* **99**:171-184.
- Jannoo N., Grivet L., David J., D'Hont A., Glaszmann J.C. (2004)** Differential chromosome pairing affinities at meiosis in polyploidy sugarcane revealed by molecular markers. *Heredity* **93**:460-467.
- Jestin C., Lodé M., Vallée P., Domin C., Falentin C., Horvais R., Coedel S., Manzanares-Dauleux M.J., Delourme R. (2011)** Association mapping of quantitative resistance for *Leptosphaeria maculans* in oilseed rape (*Brassica napus* L.). *Molecular Breeding* **27**:271-287.
- Jordan D.R. Casu R.E., Bresse P., Carroll B.C., Berding N., McIntyre C.L. (2004)** Markers associated with stalk number and suckering in sugarcane colocate with tillering and rhizomatousness QTLs in Sorghum. *Genome* **47**:988-993.
- Klett P. and Rott P. (1994)** Inoculum sources for the spread of Leaf Scald Disease of sugarcane caused by *Xanthomonas albilineans* in Guadeloupe. *Journal of Phytopathology* **142**:283-291.
- Komor E. (2011)** Susceptibility of sugarcane, plantation weeds and grain cereals to infection by *Sugarcane Yellow Leaf Virus* and selection by sugarcane breeding in Hawaii. *Eur J. Plant Pathol.* **129**:379-388.

- Lehrer A.T., Wu K.K., Komor E. (2009)** Impact of *Sugarcane Yellow Leaf Virus* on growth and sugar yield of sugarcane. *J. Gen. Plant Pathol.* **75**:288-296.
- Lockhart B.E.L. and Cronje C.P. (2000)** Yellow leaf syndrome. In : *A guide to sugarcane disease*. P. Rott, R.A. Bailey, J.C. Comstock, B.J. Croft, A.S. Saumtally (Eds.) La librairie du Cirad, Montpellier, France.
- Long A. and Langley C. (1999)** The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. *Genome research* **9**:720-731.
- Lopes S.A., Damann K.E., Grelen L.B (2001a)** *Xanthomonas albilineans* diversity and identification based one Rep-PCR fingerprints. *Current Microbiology* **42**:155-159.
- Lopes S.A., Damann K.E., Hoy J.W., Grisham M.P. (2001b).** Infectivity titration for assessing resistance to leaf scald among sugarcane cultivars. *Plant Disease* **85**:592-596.
- Lu Y., D'Hont A., Paulet F., Grivet L., Arnaud M., Glaszmann J.C. (1994)** Molecular diversity and genome structure in modern sugarcane varieties. *Euphytica* **78**:217-226.
- Maia I.G., Gonçalves M.C., Arruda P., Vega J. (2000)** Molecular evidence *Sugarcane Yellow Leaf Virus* (ScYLV) is a member of the *Luteoviridae* family. *Arch. Virol.* **145**:1009-1019.
- Marconi TG, Costa EA, Miranda HR, Mancini MC, Cardoso-Silva CB, Oliveira KM, Pinto LR, Mollinari M, Garcia AA, Souza AP. (2011)** Functional markers for gene mapping and genetic diversity studies in sugarcane. *BMC Res Notes* **4**: 264.
- Meyer J. (1989)** Histoire du sucre. *Les éditions Desjonquières, Paris*, 335pages.
- Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA. (2007)** Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res* **17**:240-248.
- Ming R., Liu S.C., Lin Y.R., Da Silva J., Wilson W., Braga D., Van Deynze A., Wenslaff T.F., Wu K.K., Moore P.H., Burnquist W., Sorrells M.E., Irvine J.E., Paterson A.H. (1998)** Detailed alignment of *Saccharum* and *Sorghum* chromosomes: Comparative organization of closely related diploid and polyploid genomes. *Genetics* **150**: 1663-1682.
- Ming R., Liu S.C., Moore P.H., Irvine J.E., Paterson A.H. (2001)** QTL analysis in a complex autopolyploid: Genetic control of sugar content in sugarcane. *Genome Res* **11**: 2075-2084.
- Ming R., Liu S.C., Bowers J.E., Moore P.H., Irvine J.E., Paterson A.H. (2002a)** Construction of a *Saccharum* consensus genetic map from two interspecific crosses. *Crop Science* **42**:570-583.
- Ming R., Wang Y.H., Draye X., Moore P.H., Irvine J.E., Paterson A.H. (2002b)** Molecular dissection of complex traits in autopolyploids: mapping QTL affecting sugar yield and related traits in sugarcane. *Theoretical and Applied Genetics* **105**:332-345.

- Ming R., Del Monte T.A., Hernandez E., Moore P.H., Irvine J.E., Paterson A.H. (2002c)** Comparative analysis of QTLs affecting plant height and flowering among closely-related diploid and polyploidy genomes. *Genome* **45**:794-803.
- Moonan F., Molina J., Mirkov E. (2000)** *Sugarcane Yellow Leaf Virus*: an emerging virus that has evolved by recombination between luteoviral and poleroviral ancestor. *Virology* **269**:156-171.
- Mukerjee S.K. (1957)** Origin and distribution of *Saccharum*. *Botanical Gazette* **119**: 55-61.
- Murray S.C.R., Rooney W.L., Hamblin W.L., Mitchell S.E. et Kresovich S. (2009)** Sweet sorghum genetic diversity and association mapping for brix and height. *The plant Genome* **2**(1):48-62.
- Myles S., Peiffer J., Brown P., Ersoz E., Zhang Z., Costich D., Buckler E. (2009)** Association mapping: critical considerations shift from genotyping to experimental design. *The plant cell* **21**:2194-2202.
- Nibouche S., Raboin L.M., Hoarau J.Y., D'Hont A., Costet L. (2011)** Quantitative trait loci for sugarcane resistance to the spotted stem borer *Chilo sacchariphagus*. *Molecular breeding*- in press - DOI: 10.1007/s11032-010-953-0.
- Offman B. (2000)** Caractérisation et analyse génétique de la résistance de la canne à sucre à *Xanthomonas albilineans*. Université de la Réunion – Thèse de doctorat.
- Oliveira K., Pinto L., Marconi T., Margarido G., Pastina M., Teixeira L., Figueira A., Ulian E., Garcia A., Pereira Souza A. (2007)** Functional integrated genetic linkage map based on EST-markers for a sugarcane (*Saccharum* spp.) commercial cross. *Molecular breeding* **20**:189-208.
- Oriol P. and Domaingue R. (2008)** Varietal selection of sugarcane. In : *Sugarcane – Montpellier, CIRAD*. 2 pages.
- Ovalle W., Comstock J., Juarez J. et Soto G. (1995)** First report of leaf scald of sugarcane (*Xanthomonas albilineans*) in Guatemala. *Plant Disease* **79**:212.
- Pan Y.B., Grisham M.P., Burner D.M. (1997)** A polymerase chain reaction protocol for the detection of *Xanthomonas albilineans*, the causal agent of sugarcane leaf scald disease. *Plant Disease* **81**:189-194.
- Pan Y.B., Burner D.M., Legendre B.L., Grisham M.P., White W.H. (2004)** An assessment of the genetic diversity within a collection of *Saccharum spontaneum* L. with RAPD-PCR. *Genetic Resources and crop evolution* **51**: 895-903.
- Panje R.R. and Babu C.N. (1960)** Studies in *Saccharum spontaneum*. Distribution and geographical association of the chromosome number. *Cytologia* **25**: 152-172.
- Patterson N., Price A.L., Reich D. (2006)** Population structure and eigenanalysis. *PLoS Genetics* **2**(12):2074-2093.
- Perrier X. and Jacquemoud-Collet J.P. (2006)** DARwin software. <http://darwin.cirad.fr/darwin>.

- Pfeffer S., Dunoyer P., Heim F., Richards K. E., Jonard G., Ziegler-Graff V. (2002)** P0 of Beet Western Yellows Virus is a suppressor of posttranscriptional gene silencing. *Journal of Virology* **76**(13): 6815-6824.
- Pfender WF, Saha MC, Johnson EA, Slabaugh MB. (2011)** Mapping with RAD (restriction-site associated DNA) markers to rapidly identify QTL for stem rust resistance in *Lolium perenne*. *Theor Appl Genet* **122**:1467-1480.
- Pieretti I., Royer M., Barbe V., Carrere S., Koebnik R., Cociancich S., Couloux A., Darrasse A., Gouzy J., Jacques M.A., Lauber E., Manceau C., Mangenot S., Poussier S., Segurens B., Szurek B., Veerdier V., Arlat M., Rott P. (2009)** The complete genome sequence of *Xanthomonas albilineans* provides new insights into the reductive genome evolution of the xylem-limited Xanthomonadaceae. *BMC Genomics* **10**:616-631
- Pinto L.R., Garcia A.A.F., Pastina M.M., Teixeira L.H.M., Bressiani J.A., Ulian E.C., Bidoia M.A.P., Souza A.P. (2010)** Analysis of genomic and functional RFLP derived markers associated with sucrose content fiber an Yield QTLs in a sugarcane (*Saccharum* spp.) commercial cross. *Euphytica* **172**:313-327.
- Piperidis G., Piperidis N., D'Hont A. (2010)** Molecular cytogenetic investigation of chromosome composition and transmission in sugarcane. *Mol. Genet. Genomics* **284**:65-73.
- Poland J., Bradbury P.J., Buckler E.S., Nelson R.J. (2011)** Genome wide nested association mapping of quantitative resistance to northern leaf blight in maize. *PNAS* **108**(17):6893-6898.
- Price S. and Daniels J. (1968)** Cytology of South Pacific sugarcane and related grasses, with special reference to Fiji. *J. Hered.* **59**:141-145.
- Price A.L., Patterson N.J., Plenge R.M., Weinblatt M.E., Shadick N.A., Reich. D. (2006)** Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* **38**:904–909.
- Pritchard JK, Stephens M, Donnelly P (2000a)** Inference of population structure using multilocus genotype data. *Genetics* **155**:945–959.
- Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000b)** Association mapping in structured populations. *American Journal of Human Genetics* **67**:170–181.
- Quesada T., Gopal V., Cumbie W.P., Eckert A.J., Wegrzyn J.L., Neale D.B., Goldfard B., Huber D.A., Casella G., Davis J.M. (2010)** Association mapping of disease resistance in natural population of Loblolly pine (*Pinus taeda* L.). *Genetics* **186**: 677-686.
- Raboin L.M. (2005)** Génétique de la résistance au charbon de la canne à sucre causé par *Ustilago scitaminea* Syd. : Caractérisation de la diversité génétique du pathogène, cartographie de QTL dans un croisement bi-parental et étude d'associations dans une population de cultivars modernes. *ENSA Montpellier, France - Thèse de doctorat*.

- Raboin L.M., Oliveira K.M., Lecunff L., Telismart H., Roques D., Butterfield M., Hoarau J.Y., D'Hont A. (2006)** Genetic mapping in the high polyploid sugarcane using a bi-parental progeny; identification of a gene controlling stalk colour and a new rust resistance gene. *Theoretical and Applied Genetics* **112**:1382–139.
- Raboin L.M., Pauquet J., Butterfield M., D'Hont A., Glaszmann J.C. (2008)** Analysis of genome-wide linkage disequilibrium in the highly polyploid sugarcane. *Theoretical and Applied Genetics* **116**:701-714.
- Rafalski A. (2002)** Applications of single nucleotide polymorphisms in crop genetics. *Current Opinion in Plant Biology* **5**:94-100.
- Ramalingam J., Vera Cruz C.M., Kukreja K., Chittoor J.M., Wu J.L., Lee S.W., Baraoidan M., George M.L., Cohen M.B., Hulbert S.H., Leach J.E., Leung H. (2003)** Candidate defense genes from rice, barley and maize and their association with qualitative and quantitative resistance in rice. *Molecular Plant-Microb Interaction* **16**:14-24.
- Rassaby L., Girard J.C., Letourmy P., Chaume J., Irely M.S., Lockhart B., Kodja H., Rott P. (2003)** Impact of *Sugarcane Yellow Leaf Virus* on sugarcane yield and juice quality in Réunion island. *European Journal of Plant Pathology* **109**: 459-466.
- Rassaby L., Girard J.C., Lemaire O., Costet L., Kodja H., Lockhart B.E.L. et Rott P. (2004)** Spread of *Sugarcane Yellow Leaf Virus* in sugarcane plants and fields on the island of Réunion. *Plant pathology* **53**:117-125.
- Reffay N., Jackson P.A., Aitken K.S., Hoarau J.Y., D'Hont A., Besse P., McIntyre C.L. (2005)** Characterization of genome regions incorporated from an important wild relative into Australian sugarcane. *Molecular Breeding* **15**:367-381.
- Remington D.L., Thornsberry J.M., Matsuoka Y., Wilson L.M., Whitt S.R., et al. (2001)** Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc. Natl. Acad. Sci. USA* **98**:11479-11484.
- Renier A., Vivien E., Cociancich S., Letourmy P., Perrier X., Rott P., Royer M. (2007)** Substrate specificity-conferring regions involved in albicidin pathotoxin biosynthesis are highly conserved within the species *Xanthomonas albilineans*. *Applied and Environmental Microbiology* **73**:5523-5530.
- Rice W.R. (1989)** Analyzing tables of statistical tests. *Evolution* **43**(1):223-225.
- Roach B.T. (1972)** Nobilization of sugarcane. *Proc.Int. Soc. of Sugarcane Technol.* **14**:206-216.
- Roach B.T. (1986)** Evaluation and breeding use of sugarcane germplasm. *Proc. Int. Soc. Sugarcane Technol.* **19**: 492-501.
- Rossi M., Araujo P.G., PAulet F., Garsmeur O., Dias V.M., Chen H., Van Sluys M.A., D'Hont A. (2003)** Genomic distribution and characterization of EST-derived resistance gene analogs (RGAs) in sugarcane. *Molecular Genetics and Genomics* **269**:406-419.

- Rowe H, Renaut S, Guggisberg A. (2011)** RAD in the realm of next-generation sequencing technologies. *Mol Ecol* **20**:3499-3502.
- Rott P., Chatenet M., Granier M., Baudin P. (1988)** L'échaudure des feuilles de la canne à sucre provoquée par *Xanthomonas albilineans* (Ashby) Down. II. Diagnostic et spectres d'hôtes de l'agent pathogène en Afrique tropicale. *L'Agronomie Tropical* **43**:236-243.
- Rott P., Abel M., Soupa D., Feldmann P. (1994)** Population dynamics of *Xanthomonas albilineans* in Sugarcane plants as determined with an Antibiotic-Resistant Mutant. *Plant disease* **78**:241-247.
- Rott P., Soupa D., Brunet Y., Feldman P., Letourmy P. (1995)** Leaf scald (*Xanthomonas albilineans*) incidence and its effect on yield in seven sugarcane cultivars in Guadeloupe. *Plant pathology* **44**:1075-1084.
- Rott P., Bailey R.A., Comstock J.C., Croft B.J. et Saumtally A.S. (2000)** A guide to sugarcane diseases. *La librairie du CIRAD, Montpellier, France*.
- Rott P. and Davis M. (2000)** Leaf Scald – in : Rott P., Bailey R.A., Comstock J.C., Croft B.J. et Saumtally A.S. (2000) A guide to sugarcane diseases. *La librairie du CIRAD, Montpellier, France*.
- Rott P., Mirkov T.E., Schenck S. et Girard J.C. (2008)** Recent advances in research on *Sugarcane Yellow Leaf Virus*, the causal agent of sugarcane yellow leaf. *Sugarcane International* **26**(3):18-27.
- Rott P., Fleites L., Marlow G., Royer M., Gabriel D.W. (2011)** Identification of new candidate pathogenicity factors in the xylem invading pathogen *Xanthomonas albilineans* by transposon mutagenesis. *Molecular Plant-Microbe Interactions* **24**(5):594-605.
- Saidou A.A., Mariac C., Luong V., Pham J.L., Bezançon G., Vigouroux Y. (2009)** Association studies identify natural variation at PhyC Linked to flowering time and morphological variation in pearl millet. *Genetics* **182**(3):899-910.
- SAS Institute (2008)** SAS OnlineDoc® 9.2. SAS Institute Inc., Cary, NC, USA.
- Scagliusi S. and Lockhart B. (2000)** Transmission, characterization, and serology of a *Luteovirus* associated with yellow leaf syndrome of sugarcane. *Phytopathology* **90**:120-124.
- Schenck S. and Lehrer A.T. (2000)** Factors affecting the transmission and spread of *Sugarcane Yellow Leaf Virus*. *Plant Disease* **84**:1085-1088.
- Shaik R., Pillay D., Pillay B. (2009)** Amplified fragment length polymorphisms reveal genetic differentiation among strains of *Xanthomonas albilineans*. *Journal of Microbiological Methods* **76**:43-51.
- Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., AcencioM., Alvarenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptisa C.S. et al. (2000)** The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature* **406**:151-157.

- Smith G., Borg Z., Lockhart B., Braithwaite K., Gibbs M. (2000)** *Sugarcane Yellow Leaf Virus*: a novel member of the Luteoviridae that probably arose by inter-specific recombination. *Journal of General virology* **81**:1865-1869.
- Solas M.T., Pinon D., Acevedo R., Fontaniella B., Legaz M.E., Vincente C. (2003)** Ultrastructural changes and production of a xanthan-like polysaccharide associated with scald of sugarcane leaves caused by *Xanthomonas albilineans*. *European Journal of Plant Pathology* **109**:351-359.
- Stevenson G.C. (1965)** Genetics and breeding of sugar cane. Longmans, *Tropical science series, Londres*, 284 pages.
- Stracke S., Prestler T., Stein N., Perovic D., Ordon F., Graner A. (2007)** Effects of introgression and recombination on haplotype structure and linkage disequilibrium surrounding a locus encoding *Bymovirus* resistance in Barley. *Genetics* **175**:805-817.
- Stracke S., Haseneyer G., Veyrieras J.B., Geiger H.H., Sauer S., Graner A., Piepho H.P. (2009)** Association mapping reveals gene action and interactions in the determination of flowering time in Barley. *Theoretical and applied Genetics* **188**:259-273.
- Terauchi H., Honda K.I., Yamagishi N., Kanematsu S., Ishiguro K., Hidaka S. (2003)** The N-terminal region of the readthrough domain is closely related to aphid vector specificity of *Soybean Dwarf Virus*. *Virology* **93**(12):1560-1564.
- Tsai C.C., Liu C.P., Chen C.T. (1992)** Application of monoclonal antibodies in the detection of *Xanthomonas albilineans*, the sugarcane leaf scald pathogen. *Plant Pathol. Bull.* **1**:159-165.
- Van Dillewijn C. (1960)** Botanique de la canne à sucre. *Editions Veenman & Zonen, Wageningen, Hollande*, 391p.
- Vega J., Scagliusi S., Ulian E. (1997)** Sugarcane Yellow Leaf Disease in Brazil: Evidence of association with a *Luteovirus*. *Plant disease* **81**:21-26.
- Victoria J.I., Avellaneda M.C., Angel J.C., Guzman M.L. (2005)** Resistance *Sugarcane Yellow Leaf Virus* in Columbia. *Proc. ISSCT* **25**:664-670.
- Viswanathan R. (2002)** Sugarcane yellow leaf syndrome in India: incidence and effects on yield parameters. *Sugar Cane International*. Sept./Oct: 17-23.
- Viswanathan R., Balamuralikrishnan M., Karuppaiah (2008)** Identification of three genotypes of *Sugarcane Yellow Leaf Virus* causing yellow leaf disease from India and their molecular characterization. *Virus Genes* **37**:368-379.
- Wang M.Q and Zhou G.H (2010)** A near-complete genome sequence of a distinct isolate of *Sugarcane Yellow Leaf Virus* from China, representing a sixth new genotype. *Virus Genes* **41**:268-272.
- Wei X., Jackson P., McIntyre C., Aitken K., Croft B. (2006)** Associations between DNA markers and resistance to diseases in sugarcane and effects of population substructure. *Theoretical and Applied Genetics* **114**:155-164.

- Wei X., Jackson P.A., Hermann S., Kilian A., Heller-Uszynska K., Deomano E. (2010)** Simultaneously accounting for population structure, genotype by environment interaction, and spatial variation in marker-trait associations in sugarcane. *Genome* **53**:973-981.
- Wilson L.M., Whitt S.R., Ibàñez A.M., Rocheford T.R., Goodman M.M., Buckler E.S. (2004)** Dissection of maize kernel composition and starch production by candidate gene association. *The plant cell* **16**:2719-2733.
- Wu K., Burnquist W., Sorrells M.E., Tew T.L., Moore P.H., Tanksley S.D. (1992)** The detection and estimation of linkage in polyploids using single-dose restriction fragments. *Theoretical and applied genetics* **83**:294-300.
- Yang X., Thannhauser T.W., Burrows M., Cox-Foster D., Gildow F., Gray S. (2008)** Coupling genetics and proteomics to identify Aphid proteins associated with vector-specific transmission of *Polyomavirus* (Luteoviridae). *Journal of virology* **82**(1):291-299.
- Yu J., Pressoir G., Briggs W.H., Bi I.V., Yamasaki M., Doebley J.F., McMullen M.D., Gaut B.S., Nielsen D.M., Holland J.B., Kresovich S., Buckler E.S. (2006)** A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* **38**(2):203-208.
- Yu J. and Buckler E.S. (2006)** Genetic association mapping and genome organization of maize. *Current Opinion in Biotechnology* **17**:155–160.
- Zeng L., Meredith W.R., Gutierrez O.A., Boykin D.L. (2009)** Identification of associations between SSR markers and fiber traits in an exotic germplasm derived from multiple crosses among *Gossypium* tetraploid species. *Theoretical and applied genetics* **119**:93-103.
- Zhao K., Aranzana M.J., Kim S., Lister C., Shindo C., Tang C., Toomajian C., Zheng H., Dean C., Marjoram P., Nordborg M. (2007)** An *Arabidopsis* example of association mapping in structured samples. *PLoS Genetics* **3**:71-82.
- Zhao K., Tung C.W., Eizenga G.C., Wright M.H., Ali M.L., Price A.H., Norton G.J., Islam M.R., Reynolds A., Mezey J., McClung A.M., Bustamante C.D., McCouch S.R. (2011)** Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat. Commun.* **2**:467(1-10).
- Zhu C., Gore M., Buckler E.S. , Yu J. (2008)** Status and prospects of association mapping in plants. *The plant Genome* **1**(1):5-20.
- Zhu C. et Yu J. (2009)** Nonmetric multidimensionnal scaling corrects for population structure in association mapping with different sample types. *Genetics* **182**:875-888.
- Zhao K, Tung CW, Eizenga GC, Wright MH, Ali ML, Price AH, Norton GJ, Islam MR, Reynolds A, Mezey J, McClung AM, Bustamante CD, McCouch SR. (2011)** Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat Commun* **2**: 467.

Site internet

FAOSTAT <http://faostat.fao.org>

Annexe 1 : Diagnostic SCYLV sur membranes de nitrocellulose par Immuno-empreinte

Principe : les particules virales étant localisées dans les cellules compagnes du phloème, elles sont transférées sur la membrane de nitrocellulose lors de l'impression. Leur présence est détectée par une double réaction immunologique. Le premier Anticorps (Ac), anti-SCYLV de lapin, est dirigé contre les sites antigéniques des particules virales. Le deuxième Ac, conjugué phosphatase alcaline anti-globuline de lapin, est dirigé contre le premier Ac. La réaction est révélée en ajoutant le substrat de l'enzyme. Ce substrat est composé d'un mélange de Nitro-blue-tétrazilium (NTB) et de 5bromo-4chloro-3indolyphosphate (BCIP). Sur la membrane NC, la réaction Ag-Ac est visible à la loupe binoculaire, par une coloration bleu foncé des cellules contaminées du phloème.

1. Prélèvement biologique :

- Prélèvement du faisceau central du limbe à environ 10 cm de la gaine foliaire.
- Empreinte sur membrane de nitrocellulose renforcé Whatmann OPTITRAN BAS 85 de 0.45 μm d'épaisseur, par application de la section transversale pendant deux secondes.

2. Blocage :

Premier bain : 30 minutes dans des bacs recouverts contenant de l'albumine bovine (SIGMA) dissoute à 3% dans du tampon TBS (8g/L de NaCl ; 2,42g/L de TRIS, qsp eau distillée, pH 7.5).

3. 1^{ère} hybridation

Deuxième bain : 90 minutes dans un bain de sérum de lapin anti-SCYLV (produit par Ben Lockhart) dilué au 1/10000 dans du tampon TBS-Tween-BSA 1% (0.05% de Tween).

4. Lavages

Trois rinçages successifs dans du tampon TBS-Tween pendant 5 minutes, en changeant le tampon entre chaque rinçage.

5. 2^{ème} hybridation

Troisième bain : 75 minutes dans un bain contenant un sérum de conjugué phosphatase alcaline anti-IgG de lapin (SIGMA) dilué au 1/10000^e dans du TBS-Tween-BSA à 1%.

6. Lavages

Trois nouveaux rinçages successifs dans le TBS-Tween

7. Révélation

- 4^{ème} bain : 20 minutes dans une solution de 5bromo-4chloro-3indoliphosphate BCIP/ NitroBlue Tetrazilum NTB (un comprimé pour 20 à 30 ml d'eau distillée).
- Arrêt de la réaction de coloration en plongeant les membranes dans de l'eau distillée.
- Séchage sur papier absorbant.

Résumé

Les cultivars modernes de canne à sucre ont un génome particulièrement complexe, polyploïde et aneuploïde. Ils dérivent de croisements impliquant un nombre réduit de clones ancestraux (*Saccharum officinarum* et *S. spontaneum*), ayant pour conséquence un déséquilibre de liaison (DL) appréciable chez ces cultivars. Ce travail vise à confirmer la faisabilité d'études d'associations exploitant le DL, afin d'identifier des allèles impliqués dans le déterminisme de la résistance à deux maladies : la feuille jaune (ac. *Sugarcane Yellow Leaf Virus*), et l'échaudure des feuilles (ac. *X. albilineans*). Deux populations de près de 200 cultivars ont été phénotypées pour la résistance à la feuille jaune, l'une à la Réunion et l'autre en Guadeloupe. Cette dernière a également été phénotypée pour la résistance à l'échaudure des feuilles. Les analyses d'associations ont été réalisées avec près de 4000 marqueurs (AFLP et DArT), avec différents modèles prenant en compte la structure de la population et les apparentements entre individus. Les marqueurs les plus robustes ont été sélectionnés à un risque d'erreur calculé sur l'ensemble du génome. Une étude de diversité a été réalisée sur un panel de clones ancestraux pour déterminer l'origine spécifique des marqueurs. Ces analyses ont permis d'identifier 23 marqueurs associés à la résistance à l'échaudure des feuilles, et 30 marqueurs associés à la résistance à la feuille jaune, présentant des R^2 individuels de 2,2 à 4%. Des sous-ensembles de 6 à 10 marqueurs expliquent 36 à 51% de la variation phénotypique. La majorité des sources de résistance et de sensibilité est issue de l'espèce sauvage *S. spontaneum*. Ce sous-génome semble fortement structurer la diversité agronomique des cultivars.

Mots clés : polyploïdie, génétique d'association, maladies systémiques, origine spécifique.

Summary

Modern sugarcane cultivars have a highly complex polyploid and aneuploid genome. They are derived from crosses involving a limited number of ancestral clones (*Saccharum officinarum* and *S. spontaneum*), resulting in strong linkage disequilibrium (LD) in these cultivars. This work aims to confirm the feasibility of LD-based association studies, to identify alleles involved in the determinism of resistance of two diseases: sugarcane yellow leaf (SCYLV) and leaf scald (*X. albilineans*). Two populations of about 200 cultivars were phenotyped for yellow leaf resistance, one in Reunion and another in Guadeloupe. The latter population was also phenotyped for leaf scald resistance. Association analyses were achieved using about 4000 markers (AFLP, DArT), with different models taking into account population structure and familial relatedness. The most robust markers were selected at a type-I error risk calculated across the whole genome. A diversity study was performed on a panel of ancestral clones in order to define the specific origin of markers. These analyses allowed the identification of 23 markers associated with leaf scald resistance, and 30 markers associated with yellow leaf resistance, exhibiting R^2 values between 2.2 to 4%. Subsets of 6 to 10 markers explained 36% to 51% of the phenotypic variation. The majority of resistance and susceptibility factors are inherited from the wild *S. spontaneum* species. This sub-genome seems to highly structure the agronomic diversity of cultivars.

Key words: polyploidy, genome-wide association, systemic disease, specific origin.